

When payment by results is a sensible approach

A proposal that senior German academics be paid on the basis of job performance has raised a storm of protest. But that should not obscure the value of this — and other — proposals for university reform.

There is nothing like an attack on the pocket to bring people out in protest. Thus an army of German university professors is now rallying against a recommendation by the University Rectors Conference (HRK) that their pay should in future depend on research and teaching performance, rather than their length of service and number of family dependants (see page 396).

The negative reaction is unsurprising, given that the recommendation would mean lower salaries for the mediocre. But a number of remarkable reasons have been put forward in the past couple of weeks to defend the traditional fixed-salary structure — even to the extent of calling as witness the constitutional right of scientific freedom. To summarize these crudely: if professors are distracted by worries about whether they will qualify for a pay increment, they will not be psychologically free to carry out their research effectively.

Such a conservative response does not serve the research community well. Academics and their universities form part of a world that increasingly demands accountability from those who are paid from the public purse. If the universities are to retain public support, they must ensure that such accountability has teeth. At present, a German professor reigns more or less unchallenged in his (rarely her) kingdom. He will receive some research money from his university, whether or not impartial experts judge his research to be worthwhile. This has not, of course, stopped the more talented self-starters from achieving research excellence — as plenty of German Nobel prizewinners demonstrate. But it has probably prevented the less talented and self-motivated from reaching a higher level of achievement than they do.

Politicians, backed by a small number of relatively enlightened academics, have campaigned for years, against a groundswell of resistance, to change the federal 'framework' law on Germany's regionally funded universities. Such changes now give the universities, among

other things, the budgetary independence that allows them to allocate more research money to higher achievers. But increasing salaries for high achievers would require another change in federal law, in order to release academic staff from the public-servant salary scales set by the federal government. The HRK cannot influence this situation directly. But hopefully the debate that it has opened up will help persuade politicians to take the next logical step to their successful modifications to the framework law.

Sadly, the issue of pay was the only one of a set of sensible efficiency-raising recommendations put forward by the HRK to have received significant attention from the grassroots academic community. In particular, the rectors' call for universities to abandon *Habilitation*, a post-doctoral qualification traditionally required to join professorial ranks, was greeted with a resounding silence.

The HRK's call for its abandonment is only the latest in a long succession of such calls; but is no less correct for that. Hundreds of young researchers are asked to spend valuable time 'proving' through research experience in their professor's laboratory, through the writing of a further thesis and through years of supporting their professor's teaching activities, that they are suitable for consideration as professors in German universities. This method serves only to extend training time, usually until at least the age of 40; a bundle of key research papers and standard evidence of teaching experience would suffice just as well as proof of worthiness, as it does in other countries.

Any proposal by a group of employers whose effect would be to reduce the living standards of some of their employees — even though these will remain relatively high — deserves tough scrutiny. But the rectors' recommendations need to be addressed in a positive spirit if Germany's universities are to remain internationally competitive, and thus worthy of support from the German taxpayer. □

Will we see their like again?

Departing directors at the top US physics laboratories leave a discipline in search of inspired leadership.

When Burt Richter had to face Dana Rohrabacher — the abrasively right-wing congressman who had just taken on the chair of an influential committee with oversight of scientific programmes at the US Department of Energy — he found a quick way to connect with a politician more naturally at home with the National Rifle Association than with the American Physical Society, of which Richter happened to be president at the time. Richter invited Rohrabacher round to try out his pistol collection. "I wanted him to know that not all physicists are wimps," he later explained.

Such resourcefulness will be sorely missed when Richter steps down next August as director of the Stanford Linear Accelerator Center in California (see page 397). Indeed, the exit of the Nobel-prizewinning physicist will leave many scientists at the Department of Energy laboratories wondering where they will see his like again. For coming on top of last year's unfortunate departure of Nick

Samios (another blunt New Yorker) from the Brookhaven National Laboratory, New York, and next year's retirement of John Peoples from Fermilab in Illinois, Richter's departure leaves the US high-energy physics community grappling for fresh leadership at a time when its political support is far from assured.

It would be short-sighted to observe the stature of these men and conclude them to be irreplaceable. Each in turn, after all, stepped comfortably enough into the shoes of a previous generation of physicists that had lifted its immense national prestige directly from its role in the Manhattan project. But these days, it is only the weapons laboratories at Los Alamos, Sandia and Livermore that retain such prestige and the funding that accompanies it. The non-weapons laboratories stand alone, and relatively exposed. A new generation of leaders is now sorely needed to win friends, influence people and lead US particle physics into the twenty-first century. □



Iceland poised to sell exclusive rights to national health data

[LONDON] Critics are hoping that a last-minute glitch will prevent Iceland passing a controversial law that would give a private company the exclusive rights to an electronic database of the nation's health records.

The government-sponsored legislation would give an Icelandic company a 12-year licence to create and market a database combining genetic, healthcare and genealogical information on every person in the country.

Although scientists have withdrawn their criticisms after the bill was modified, the proposed law has outraged opposition parties and medical doctors. Critics are now pinning their hopes on the slim possibility that the bill could be dropped under the weight of legislation waiting to be passed in the current session of parliament.

The database will heavily encrypt the identity of individuals, and is designed to provide researchers with information on groups of possibly ten. It will be regulated by four different committees appointed by the government, including a data protection committee and an ethics committee.

Patients will be asked for their consent if genetic information is to be included. But health and family data will be added to the database automatically from hospital and doctors' records, although individuals can be removed from the database if they make a written request.

The bill will almost certainly become law if it is put to a vote in parliament before Christmas. The outcome is effectively guaranteed, as the ruling centre-right coalition government has a comfortable majority in the 63-seat parliament.

Few recent government proposals have caused as much debate in Iceland as the database bill. It receives regular media coverage, and opinion polls have shown substantial but decreasing public support for the idea.

Critics include the Icelandic Medical Association (IMA), Mannvernd, a newly formed association of scientists and medical researchers opposed to the bill, and even Ernir Snorrason, a co-founder of the Reykjavik-based genomics company deCODE Genetics, which is most likely to win the licence for the database.

But the country's nurses and the Research Council of Iceland, both former critics, have reconsidered after government changes to earlier drafts of the bill — particularly disallowing the use of data on individuals and allowing patients to leave the database.

Vilhjalmur Ludviksson, the research council's director, says that the database provides a rare opportunity for Iceland's



Gene pool: Iceland's genetically homogenous population attracts geneticists and medical researchers.

researchers to become involved in some exciting science, which the government — whose annual research budget amounts to US\$120 million — would never be able to afford. But the IMA says that the database's drawbacks outweigh any potential advantages.

Patient privacy is a key concern. Ross Anderson, an adviser on computer security to the British Medical Association asked by the IMA to assess Iceland's proposals, says that no country has so far tried to set up a centralized database of this type.

This, he says, is partly because if all the information is in one place, it is easier to decode, regardless of security measures. Anderson says it is better to link separate databases than to create a central facility.

The main criticism, however, is of giving exclusive control of a database of all Icelanders to a single company. Indeed, perhaps the most controversial aspect of the bill is that was proposed by deCODE Genetics.

Last year, the company entered into a US\$200 million agreement with Hoffmann La Roche, the Swiss pharmaceutical firm, to use separate information on Icelandic patients in research on 12 common diseases. The database bill is the brainchild of its chief executive, Kari Stefansson, and the company plans to invest US\$170 million in the project. There are no other known contenders.

Both the government and Stefansson make no secret of either the origins of the idea or the potential commercial benefits to the company from marketing access to Iceland's healthcare records.

The government, however, says that the main aim of the database is to give scientists access to Iceland's coveted health records on its genetically homogenous population.

It also argues that the database will bring

many benefits to Icelandic science, as the bill stipulates that a foreign company wanting access to the records will need to develop its proposals in collaboration with researchers in Iceland. It also specifies that the database cannot be sold overseas.

The government says that, while undoubtedly desirable, public sector investment in a database is beyond its reach, and private sector finance is needed. It adds that, without the carrot of an exclusive licence, no company would be willing to shoulder the required investment.

The government also claims that one of the advantages of a privately funded database is that the licence-holder will have to pay to manage, monitor and regulate the database, and to ensure that patient privacy is maintained. All these functions, however, will be independent of the licence-holder, and be monitored by the committees.

Despite these assurances, the IMA says that the licence-holder will retain considerable influence. deCODE Genetics is lobbying to sit on most committees, on the grounds that its interests need to be represented as the likely licence-holder.

The licence-holder will decide which companies to allow access to the data. It will also be allowed to veto use of the database by Icelandic scientists if their proposed research is judged to threaten its commercial interests.

Stefansson says that without such a veto, there is a risk that foreign multinational companies would be tempted to secretly use local scientists for commercial work. The government wants to charge local scientists less than foreign researchers. But it is uncertain whether this is allowed under the terms of the European Economic Area, of which Iceland is a member.

Ehsan Masood

FRANK SPOONER PICTURES

Germany's rectors call for pay by results

[MUNICH] A proposal by German university chiefs that the pay of academics should be based on performance has caused controversy in the academic community.

Such a system would promote competition between both academics and universities, says the German University Rectors Conference (Hochschulrektorenkonferenz; HRK). It would also allow universities to offer industrial-level salaries to attract particularly talented scientists.

But the German Association of Universities, a lobby organization representing 17,000 professors, has reacted angrily to the proposal. Its president, Hartmut Schiedermaier, describes the suggestion as "incredible" and directly against the interests of professors, as it could dramatically reduce the income of the next generation of academics.

The rectors have recommended that current increments related to age and family size should be eliminated. Instead the basic salary should be roughly equivalent to a middle grade of professor, currently around DM8,700 (US\$5,100) a month. Increments would be calculated on the basis of research and teaching performance.

In presenting the proposal at the HRK

general assembly last month, Klaus Landfried, president of the organization, admitted that, if it is adopted, "many people will start off with less money than today".

Starting salaries should be negotiated between universities and individual candidates, says Landfried. Universities would then be able to use their personnel budgets flexibly, rather than being obliged to comply with centrally fixed salary scales.

The proposal is supported by the new Social Democrat minister for education and research, Edelgard Bulmahn. In fact, such a pay reform is part of the new red-green government's coalition agreement. "Extensive modernization of employers' rights is of fundamental significance for the future of our universities," Bulmahn said in her inaugural speech to the German parliament last month.

But Schiedermaier, speaking for the German Association of Universities, says that the proposals are particularly unfair because university teachers have had to cope with many more students in the past two decades, yet have still managed to maintain high research standards.

Moreover, he strongly objects to the



HRK's suggestion that the same basic salary should be applied to teachers at Fachhochschule, the vocationally orientated higher-education institutes which have lower qualification requirements than universities.

Some rectors agree. The proposal "has triggered considerable anger among the professors at our university", says Jens Peter Meincke, a lawyer and rector of the University of Cologne.

The HRK also proposes abolishing civil servant (*Beamter*) status, equivalent to tenure, for professors. But critics of this suggestion argue that tenure is a necessary incentive for scientists whose early careers, based on short-term contracts, are difficult, poorly paid and risky. "The *Beamter* status has a long tradition, and the security it offers is important to ensure the independence of professors," says Meincke.

The German Association of Universities is also concerned whether research and teaching performance can be assessed fairly. The HRK says assessment should be based on reliable indicators such as number of research publications, amount of external research funds, prizes and awards, number of examinations set and results of teaching evaluation by students.

But the association says that it is wary of giving control of salary increments exclusively to deans or rectors. They say that although such criteria are objective, additional subjective factors would always play a role in evaluations.

The salary debate has overshadowed other recommendations in the HRK's proposal. These include the call for universities to abandon *Habilitation*, the high-level but time-consuming postdoctoral qualification officially required to become a professor.

The rectors also call for more flexibility in the career structure of technical staff, enabling the best technicians to be rewarded financially. These recommendations have prompted no response from the academic community.

Tilmann Kiessling

Software fix for Galileo may salvage data

[WASHINGTON] US space engineers are working to correct minor problems with the Galileo spacecraft that have ruined observations of Jupiter's icy moon Europa during two of the probe's last three fly-bys. The last of Galileo's close encounters with Europa is scheduled for February.

During the most recent fly-by on 22 November, the spacecraft unexpectedly went into a protective 'safe' mode just before passing the moon at a distance of 2,300 kilometres. This prevented it from taking photographs and gathering data.

Galileo project scientist Torrence Johnson of the Jet Propulsion Laboratory in Pasadena, California, says the most likely culprit is increased radiation when the spacecraft's highly elliptical orbit brings it close to the giant planet. Galileo, which left Earth in 1989 and arrived at Jupiter six years later, has already taken 150 per cent of the radiation dose for which it was designed.

Having completed its primary mission last December, the spacecraft was given a two-year extension to gather more data on Europa, study the torus of charged particles associated with Jupiter's satellite Io, and make two close fly-bys of that volcanic moon next autumn.

Europa is a high-priority target for the US space agency NASA, which plans to send a follow-on mission early in the next decade

to help decide the question of whether an ocean lies beneath the satellite's icy crust. The problems with Galileo mean that some of the moon's surface will not be photographed at high resolution as planned.

But Johnson says some high-priority observations may be made by rescheduling targets to be viewed during the final fly-by, as was done after a short-circuit from a piece of space debris led to the loss of Europa data from an encounter last July.

Project engineers are working on a software fix that would in effect tell Galileo to ignore circumstances that trigger the protective mode during close encounters. This is not likely to be ready by February, however, so the last fly-by may meet the same problem.

The strategy is also risky because it is akin to disconnecting an alarm. But the extended mission was always planned to be riskier, says Johnson.

Hoping that the spacecraft survives the high-radiation environment around Io next year, project scientists are drawing up a proposal to NASA for another one-year extension. That would allow Galileo to make coordinated observations with the Saturn-bound Cassini spacecraft when it flies through the Jupiter system in December 2000.

Tony Reichhardt

Indian farmers burn transgenic cotton crop in field trial

[NEW DELHI] Farmers protesting against genetically engineered crops last week destroyed a plot in the Raichur district of India where *Bt* cotton produced by the US company Monsanto was undergoing field trial.

Some 60 members of Karnataka State Farmers' Association, led by its president, M. D. Nanjundaswamy, uprooted the plants and burnt them. They described their action as "a message to all those who have invested in Monsanto to take their money and get out".

The plot is one of the 40 locations in India where the *Bt* cotton, which is resistant to bollworm, is being tested by Maharashtra Hybrid Seed Company, in which Monsanto has 26 per cent shares. Permission for the trials was given by the Department of Biotechnology (DBT), and the trials were expected to be concluded by end of 1999.

A DBT official, P. K. Ghosh, says destruction of the crop was unwarranted as the trial plots were being monitored by his department to check for escape of pollen. "The trials posed no biosafety concern and toxicity tests in animals showed the *Bt* cotton is perfectly safe and did not cause allergic reactions," according to Ghosh, who says that DBT was going to clear the crop for commercial production (see *Nature* 388, 817; 1997).

Four more crops — potato, tomato, cauliflower and tobacco — carrying the *Bt* gene are also undergoing field trials, and a high-yielding transgenic mustard is being grown in 20 locations. According to Ghosh, all these crops will be ready for large-scale trials next year, subject to government clearance.

But critics such as Vandana Shiva, president of the Research Foundation for Science, Technology and Ecology, are calling for an immediate halt to Monsanto's trials and a five-year moratorium on the commercialization of genetically engineered crops while adequate ecological and regulatory frameworks are developed.

K.S. Jayaraman

Richter quits as director of Stanford accelerator lab

[WASHINGTON] Burt Richter, one of the most influential — and certainly the most colourful — of the directors of the Department of Energy laboratories, is to step down next August as director of the Stanford Linear Accelerator Center (SLAC) in California. He will remain at SLAC, however, working on research and science policy issues.

Richter, who won the 1976 Nobel Prize for Physics for his work at Stanford on developing a particle collider that led to the discovery of the J/ψ particle, became director of SLAC in 1984 when Wolfgang Panofsky retired. Even though he has been director of SLAC for 14 years, the unexpected departure of such a highly motivated scientific leader created a maelstrom of rumours, in California and Washington, about its true rationale. Stories that Richter had been forced out were dismissed by his former deputy, Sidney Drell, as "a load of nonsense".

Richter leaves SLAC in good shape, with the B-Factory, a new facility for producing B-mesons, having been inaugurated in October, and an upgrade under way at the laboratory's synchrotron light source. But his departure leaves others to pursue his dream of building the Next Linear Collider (NLC), which many see as the world's next major particle physics accelerator, in California.

Richter had recently learned that the start of the conceptual design phase of the NLC will not be included in the Department of Energy's (DOE's) budget proposal for the year 2000. But he denies that his decision to step down was affected by this or by the most recent of his arguments with his superiors at the DOE — a row with Martha Krebs, the assistant energy secretary, and her boss at the time, Federico Peña, over the allocation of money for science education between the laboratories and DOE headquarters.

"I don't have any problem with the DOE," Richter told *Nature* last week. He added that



Richter: denies rift with Washington.

he had been thinking of stepping aside for a couple of years, and told Gerhard Casper, the president of Stanford University, of his decision in the summer.

Richter points out that construction of the NLC is not due to begin until 2003 at the earliest. "The next generation should take care of it now," he says, adding that he will still be director for another nine months, and intends to keep pursuing research to support the NLC.

Richter has vigorously pursued the idea of building the NLC somewhere in California — SLAC's site is too small to accommodate it — in close collaboration with Japan and other nations. But, according to one government official, Richter has not done enough to engage other US laboratories in the project, and the DOE would like to see more laboratories involved, partly to broaden its political support.

Despite his single-minded pursuit of SLAC's interests, Richter will be missed by other laboratory directors, who admired his chutzpah.

"He has been one of the best and most clear-thinking of the laboratory directors," says Bill Madia, director of the Pacific Northwest National Laboratory in Washington state. "There's no sugar-coating with Burt — he calls it as he sees it."

Richter says that, although he has enjoyed supervising research at SLAC, other aspects of the job, which involved dealing with the world outside, "have been interesting but stressful. Your enthusiasm for it reaches a maximum and then begins to get exhausted. Then it is time to step down."

Colin Macilwain

Moscow scientists reject funding deal and plan more protests

[MOSCOW] Scientists in Moscow have vowed to continue protesting at the lack of adequate funding for research, rejecting an agreement reached last month between the Russian Committee of Scientific Collectives (RCSC) and the deputy prime minister Vladimir Bulgak.

Although the government has, in line with the agreement, paid the scientists their salaries for November, a meeting of the trade unions of the Moscow branch of the Russian Academy of Sciences (RAS) issued a statement saying that they felt "free to break

the achieved agreements" (see *Nature* 396, 208; 1998).

The researchers insist that, before the end of the year, the cabinet should transfer 5 billion rubles (US\$28 million) to scientific organizations, including 2.9 billion rubles in salaries. These figures were approved earlier this year by the previous government, which was dissolved in August.

But these figures are significantly higher than those in the agreement signed by the RCSC, which is the Russian trade union of scientific workers. This gives only 1.9 billion

rubles, including 1.8 billion in scientists' salaries. The cabinet has promised to pay another 1.1 billion rubles to scientists in January and February next year, and the remaining 2 billion will be provided in exemptions from utility charges.

The Moscow RAS trade unions say that they cannot accept a situation in which there is virtually no money for scientific work itself, such as equipment, reagents, conferences and seminars. They are seeking the support of other regional branches for their position.

Carl Levitin

Anger over French synchrotron decision

[PARIS] French scientists are challenging a government decision not to earmark any funds to build a new FF1.4 billion (US\$244 million), 2.15-GeV synchrotron radiation facility, Soleil, in the 1999 science budget.

Many researchers had hoped that construction would be given the go-ahead at an interministerial committee meeting in January. But Claude Allègre, the minister of research, has described the facility in a parliamentary debate as costly and only of interest to "about 40 researchers". Referring to foreign access to the facility, Allègre said that building the machine would be a financial subsidy "to British research".

The minister also claimed that much of the impetus for construction originated in rivalry between the regions seeking to host Soleil. Finally, Allègre claimed that French scientists would not suffer because they could continue to use other existing third-generation synchrotron sources (see *Nature* 395, 831; 1998).

Several directors of synchrotron facilities, however, have signed a joint letter saying that it is impossible for them to accommodate "a scientific community as large and diversified as the French". Signatories of the letter include Nils Mårtensson of MAX-lab at Lund in Sweden, Herbert Moser of ANKA at Karlsruhe in Germany and Jochen Schneider of HASYLAB in Hamburg.



The LURE source at Orsay: already heavily used.

Yves Petroff, director-general of the European Synchrotron Radiation Facility at Grenoble, says he does not know "where Allègre got his figures from, but they are so ridiculous it is not even worth replying". The figure of 40 refers not to the number of researchers but to the number of beam lines that Soleil will offer, says Petroff, adding that more than 2,000 researchers use the second-generation light source LURE at Orsay.

The candidate host regions have asked for the decision on whether to build Soleil to be

independent of that of where to site it. Arguing that "the decision to construct Soleil is a top priority", they have also agreed that the host region will contribute to funding the light source and the planned beam lines during the initial programme, while the other regions will pay to build additional beam lines specific to the activities of their researchers. Allègre has been aware of this agreement for several months, as a letter signed by the six main scientific backers of the project was sent to him in July. No reply has yet been received.

François Willeumier, chair of the committee 'Soleil for the Ile-de-France', points out that this arrangement is already working for two beam lines at LURE. Similar agreements exist with Spain and Switzerland.

Several researchers at the Centre National de la Recherche Scientifique (CNRS) say that Allègre's desire to encourage interdisciplinary research and create a centre for genome research (Génopole) is inconsistent with his refusal to build a machine that is essential for the future of genomic research.

Without Soleil, say Christian Le Brun and Jean-Pierre Couture, directors of research at the CNRS, "France will become an underdeveloped country" at a time when the major European nations are equipping themselves, as in the United Kingdom, or have already done so, as in Germany.

Eric Glover

Reprieve on the cards for reform to Italian research council

[MUNICH] Italy's new government, which came to power in October, is expected to modify proposals for the reform of the National Research Council (CNR), which is responsible for more than 300 research institutes and centres.

As a result, the council could maintain part of its grant-giving role — this would have been removed under the earlier version of the reforms — and keep its scientific advisory committees under its own wing.

The earlier reforms, intended to increase the efficiency with which Italian science is organized, were included in a decree which failed to receive complete approval before the collapse of the government of former prime minister Romano Prodi. But the government did approve a measure to dissolve the current organization of the CNR at the end of this year.

Decrees covering organizational reforms to the Italian Space Agency and the energy and environment agency have been approved by parliament in the past few weeks. But the CNR decree was held back for review by the new research minister, Ortensio Zecchino (see *Nature* 395, 827; 1998).

Zecchino, a member of the Catholic PPI

party (Partito popolare italiano), has close political links with Lucio Bianco, the president of CNR. Bianco had felt that some aspects of the original decree weakened the influence and autonomy of the CNR, partly because members of its committees would be appointed by the research ministry (see *Nature* 395, 203; 1998).

Bianco put forward a series of modifications to the decree which were presented to parliament last week. First, he suggested that the planned discipline-orientated *consigli nazionale* (scientific advisory committees), which will serve the new government-level committee of experts for scientific policy, should be within the CNR, rather than in the research ministry as planned. He suggested that they should also be responsible for advising the CNR board.

When the current CNR organization is dissolved at the end of this year, the research council's 15 scientific advisory committees, which have been responsible for allocating research funds and advising the government on scientific issues, will disappear.

By tapping directly into the new government scientific advisory structure, Bianco hopes to restore some of the political

influence the CNR has traditionally enjoyed. But some scientists are worried that if the CNR uses the *consigli nazionali* as its advisory committees, rather than creating its own, it could mean a continuation of the influence of powerful university professors, known as *baroni*, over CNR policy.

Bianco says "there will be no return to the power of the *baroni*" because the committees will be advisory only and will not directly influence the allocation of funds. But many CNR scientists would still prefer their own advisory committees, where they could be represented in greater numbers, even if these did not have the direct ear of the government.

Bianco had been unhappy that the grant-giving role of the research council was to be transferred to other organizations, including the research ministry (see *Nature* 394, 712; 1998). His second suggestion returns to the CNR the responsibility for allocating Italy's large research grants for politically defined strategic projects, as well as small grants to allow young researchers to test new ideas.

The new decree has now to be approved by the government.

Alison Abbott

Rescued satellite to get more managers

[WASHINGTON] Officials in charge of the newly recovered Solar and Heliospheric Observatory (SOHO) satellite are proposing that an extra layer of US–European management be added to guard against mishaps of the kind that nearly led to the satellite's loss last summer.

The proposal has been worked out by space science managers at the European Space Agency (ESA) and NASA's Goddard Space Flight Center. It would reinstate a SOHO programme office at Goddard that was due to be scrapped this autumn as the mission moved into a leaner, extended phase beyond its original two-year lifetime.

Management of the \$1 billion satellite was to have merged with that of other projects in the International Solar-Terrestrial Physics (ISTP) programme to save money. The proposed reinstated SOHO office would have minimal staff: a NASA programme manager and an ESA deputy, who would ensure that ground procedures receive closer scrutiny — particularly when procedures are changed.

Although Goddard handles ground operations for SOHO, ESA has final responsibility for the spacecraft's safety. The European deputy would therefore assume command in the event that SOHO's health was in jeopardy.

The proposal was presented this week to a NASA–ESA review board set up to decide whether SOHO is ready to resume normal scientific operation. The staffing plan will have to be approved by both space agencies.

A NASA–ESA investigation board concluded last August that a string of procedural and management errors led to the June mishap, during which the satellite's solar panels turned away from the Sun and drained nearly all the power from the batteries.

Among the problems identified by the board were inadequate review of changes to spacecraft command procedures, and an overloaded ground crew that was working too fast with too few people. The board called on the agencies to review staffing levels, and to add staff as required.

SOHO was brought back to life after heroic efforts by ESA and NASA teams working with engineers from satellite manufacturer Matra Marconi Space (see *Nature* 394, 5 & 395, 313; 1998). All 12 of the satellite's scientific instruments have now been restored to full health, although one coronagraph, for an instrument that tracks eruptions from the Sun, is still being tested.

John Credland, head of ESA's scientific projects department, says that his agency has spent several million dollars on the recovery and that, despite the near-miraculous comeback, neither ESA nor NASA want to have to repeat the experience. "If we just let things



SOHO comeback: satellite is back in operation after it narrowly missed being lost in space.

revert to how they were before, we would not be responsible," he says.

But Credland says the incident should not be taken as an indictment of NASA's much-touted "faster, better, cheaper" approach. Mario Acuña, the ISTP project scientist at Goddard, says it is not even clear that extra staff would help, rather than improved train-

ing for contract workers at ground control. Both men say that cuts to mission operations are desirable and practical.

Yet the satellite's brush with disaster may serve as a warning that there are limits to the extent to which missions can be squeezed. NASA and Lockheed-Martin, the agency's new contractor for spacecraft ground operations across the board, are dedicated to cutting ground crews as much as possible to reduce the cost of space science missions.

According to Credland, the SOHO team started with 24 people, was planned to shrink to around 16, and is now likely to end up somewhere in between. The staff for the overall ISTP started at 200, but was planned to drop to 65 or 70 during the extended mission funded by NASA to continue observations during the forthcoming solar maximum. Acuña admits that the target staffing level is probably too low.

He says this week's readiness review is something of a formality, as SOHO is already back to full operation. **Tony Reichhardt**

Uruguay responds to scientists' protest

[LONDON] Reacting to pressure from the scientific community, the government of Uruguay has promised to set up a fund to support basic research, with an initial budget of US\$1 million. It has also agreed to provide support for the only non-university research institute in the country, the Instituto Clemente Estable in Montevideo.

Both moves follow a protest letter sent by scientists in October to the country's president, Julio Maria Sanguinetti, complaining that the government's apparent lack of interest in supporting research was threatening to lead to the collapse of the science base.

Scientists argue that, since the restoration of democracy in Uruguay in 1985, the country has managed to develop relatively strong activity in science and technology, despite the small size of the scientific community (Uruguay spent only 0.15 per cent of its gross national product on research and development in 1995, compared to 0.9 per cent in Brazil).

But many fear that this achievement is now at risk as the government focuses on other funding priorities. They point out that the government has failed to produce a promised five-year plan for science and technology, even though its term of office only has one more year to run.

The recent protest was triggered by difficulties at the Instituto Clemente Estable, founded by a former pupil of the Spanish neuroanatomist Santiago Ramon y Cajal. The private institute, one of the

leading biological research labs in South America, had been threatened with closure as a result of a decision by the Congress to block its public funding.

In their letter to Sanguinetti, seeking support for the institute, the scientists also pointed out that salaries for researchers in Uruguay are considerably lower than in neighbouring countries, such as Brazil and Argentina. Andres Lalane, the president of the science council CONICYT, says poor living and working conditions are driving scientists abroad or into other careers.

It is partly to address this situation that the government has now agreed to set up a National Research Fund. This will be run by the president of the University of the Republic, the president of CONICYT, and the Ministry of Education and Culture.

The future of CONICYT itself, however, remains uncertain, despite demands from the scientists that its role should be more clearly defined. CONICYT is the main agency for promoting science in Uruguay.

Since its creation 30 years ago, the organization has operated largely independently of the government, bringing together researchers from the public and private sectors. There are now fears that the government plans to reduce its autonomy and the transparency with which it operates.

Despite such concerns, Uruguayan scientists say they are optimistic that the new initiatives are hopeful signs that a more productive dialogue may be developing with the government. **Andrea Kauffmann-Zeh**

Research heads seek European platform

[STRASBOURG] The science ministers of Austria and Germany are supporting a bid by the European Science Foundation (ESF) to become a formal adviser to the European Commission on research issues.

At the same time, however, tensions have arisen between the ESF and the heads of the national research councils of the European Union (EU) — who meet regularly as a group known as Eurohorcs — over the amount of direct influence that the research councils should, through Eurohorcs, have within the ESF.

ESF's main activities include supporting European research networks and conferences and coordinating reports on issues such as the future needs of European researchers for synchrotron radiation or neutron sources (see *Nature* 396, 4; 1998).

Under pressure from its 62 member organizations, from 21 European states, to expand its remit to include activities not undertaken in the European Commission's framework research programmes, the ESF has in recent years developed interests in research policy.

Two years ago, for example, it presented a detailed report to the commission on a proposed scientific agenda for Europe, prepared for the debate on the content of the fifth Framework research programme (*Nature* 382, 8; 1996).

Casper Einem, the Austrian research minister, told the ESF annual assembly in Strasbourg last week that its unique constitution as a federation of academies of sciences and research councils made the foundation a suitable body to "take up European tasks such as evaluating European [research] programmes and being consulted on all science policy questions".

Reinhard Grunwald, secretary-general of the Deutsche Forschungsgemeinschaft, the German university grant-giving body, told the assembly that the new German minister of research, Edelgard Bulmahn, also believed that the ESF should have a formal advisory role to the EU.

Grunwald said that this might be done through the European Research Forum, the commission's new top-level research advisory committee (see *Nature* 393, 502; 1998).



Brooks: wants role in Strasbourg body.

Bulmahn will put this on the agenda of the regular EU research ministers' meeting when Germany takes on the EU presidency in January.

Meanwhile, however, questions are being raised about the extent to which the foundation can simultaneously serve the interests of its "intellectual stakeholders"—the grassroots scientific community — and its "financial stakeholders", namely its member organizations, including in particular those belonging to Eurohorcs.

In contrast to the ESF, Eurohorcs has no secretariat or executive. Indeed, it takes pride in the informality of its loose organization. The group meets twice a year, primarily to discuss shared problems.

Despite the lack of a formal structure, Eurohorcs is keen to have a stronger voice in European-level science policy decisions. Aware that its councils provide most of the ESF's funding, it has been pressing the ESF to act in a more defined way as the executive — and voice — of Eurohorcs.

At last week's assembly meeting, for example, Richard Brook, head of the UK Engineering and Physical Sciences Research Council, and chair of Eurohorcs, called for a clear relationship with Eurohorcs to be written into the ESF statutes. "We need defined links that everyone can understand," he said.

But ESF general secretary Enric Banda points out that such a formal relationship can only be incorporated into the foundation's statutes if Eurohorcs were to become a legal entity.

Not all Eurohorcs members want this. "We should remain an informal organization," Grunwald told the assembly. Unlike Brook, Grunwald wants the relationship between the ESF and Eurohorcs to "be maintained more through mutual trust than institutional links".

Banda is frustrated by what he sees as a lack of clarity in the demands of some Eurohorcs members and their vague but frequently articulated mistrust of the ESF.

Banda says he would welcome closer cooperation with Eurohorcs, "But the ball is their court". He points out that Eurohorcs "has it in its power to take over the ESF by dominating the executive council, which is made up of representatives of member organizations".

Alison Abbott

Australia's basic research faces shake-up

[SYDNEY] A plan for the radical transformation of the funding of basic research in Australia has raised a storm of protest after extracts from a confidential government document were leaked to the press.

Complaints focus on the proposed ending of the Australian Research Council's competitive, peer-reviewed funding system. Instead, research in science and the humanities would be supported by block grants to institutions.

After a formula for dividing up funds between institutions has been settled — by performance criteria which many fear will favour the larger, older universities — the ARC (current budget A\$445 million; US\$280 million) would see its influential role contract to a purely advisory one.

Researchers are angry that draft details have not been circulated publicly for comment, claiming that the education department has been feeding stories to the media to gauge the strength of feeling.

The government has given no indication of the reasoning behind these changes. Its apparent unwillingness to discuss the proposals openly is mirrored by the silence of Vicki Sara, chair of the ARC, who had been expected to release recommendations of an external review and a strategic plan for the ARC in the near future (see *Nature* 389, 220; 1997).

Sara had argued that the council's status

should change from a that of a branch within the education department to a statutory authority, but such enhanced independence is opposed by the conservative government that was re-elected in October.

According to leaked reports, the Australian National University — currently treated separately from other universities — may be absorbed into a single national scheme. If so, its Institute of Advanced Studies would lose its unique ability to mount long-term programmes with block funding — IAS staff are ineligible for ARC grants (see *Nature* 382, 484; 1996).

Science and university leaders are lobbying the education minister, David Kemp. The sharpest reaction has come from the Federation of Australian Scientific and Technological Societies, whose president, Peter Cullen, wrote to Kemp last week expressing "deep concern".

Cullen urged Kemp to maintain government support for basic research by keeping the current scheme. He says that the federation, which represents over 40 specialist societies, "doubts that an internal university reviewing process can meet the [international] standard developed by the ARC" and believes funds "should be focussed only on the most excellent of proposals". He urges Kemp to "put an end to these damaging and ill-considered ideas".

Peter Pockley

Japan split over US aid for spy satellites

[TOKYO] Plans for four reconnaissance satellites have become a bone of contention between the Japanese government and the Liberal Democratic Party (LDP), the largest party in the coalition government, centring on whether Japan should ask the United States for technical help.

The satellites are intended to improve Japan's ability to gather security information — a need highlighted by a supposed missile test by North Korea in August — and to monitor natural disasters.

The government argues that such satellites should be developed at home so as to avoid the tensions that could arise with other Asian nations if technical cooperation meant that bilateral security agreements with the United States had to be renegotiated.

But the LDP is calling strongly for US assistance on the project, claiming that Japan lacks the technology to develop and launch the satellites by its projected date.

The decision to build the satellites comes despite a resolution on the non-militarization of space passed by the Diet (Japan's parliament) in 1969. The government decided in 1985 that the resolution still allows the use of satellites for defence purposes, provided that they work in the same way as those being used for non-military purposes.

The decision to launch was announced to the cabinet last month by Hiromu Nonaka, the chief cabinet secretary. It comes in response to the firing of a North Korean missile on 31 August which crossed mainland Japan before falling into the Pacific Ocean.

As a result, the Japanese government froze various aid packages to North Korea. Although North Korea claimed that the object was a 'singing satellite' for broadcasting patriotic songs, Japan claimed it was a three-stage ballistic missile (see *Nature* 395, 312; 1998).

The US State Department has since concluded that the suspected missile was a small satellite that failed to reach orbit. But the Japanese government, which conducted its own investigation, continues to claim that it was a ballistic missile which seriously threatened national security.

The government is setting up an inter-ministerial committee with representatives from the Science and Technology Agency, the Ministry of International Trade and Industry, the Defense Agency and the Ministry of Foreign Affairs to promote the development of the satellites.

According to Nonaka, the satellite project, which will be completed by the end of the financial year 2002, will be funded by the government's third supplementary budget (see *Nature* 396, 205; 1998).

The initial stage of the project will receive around 10 billion yen (US\$83 million) from

the supplementary budget, and the total cost is expected to be ¥150 billion.

According to the government's plans, the four satellites, two of which would carry radar sensors, will be developed domestically, using technology based on satellites already in use or under development.

The National Space Development Agency (NASDA) is expected to take a central role in the satellites' development, with a proposed launch by March 2003. In particular, it is hoped that the satellites will be based on NASDA's Advanced Land Observation Satellite (ALOS), scheduled for launch in 2003 to carry out mapping, Earth resources surveys and disaster observation.

It had been speculated that ALOS itself would take on reconnaissance activities, but NASDA denies that it would be more than a technological model for the satellites. "ALOS is not going to turn into a reconnaissance satellite, and it will carry out activities in accordance with its original plan," says a NASDA spokesman.

The spokesman adds that, while the government plans to develop the satellites using Japanese technology, "NASDA has not ruled out the possibility of gaining technological assistance from other countries".

The LDP is already exploring the possibility of buying equipment and obtaining technological cooperation from private companies and government organizations in the United States in order to meet the proposed launch schedule.

Its working group on satellite development, which recently visited the United States to discuss such cooperation, says the government's project is not realistic using domestic technology.

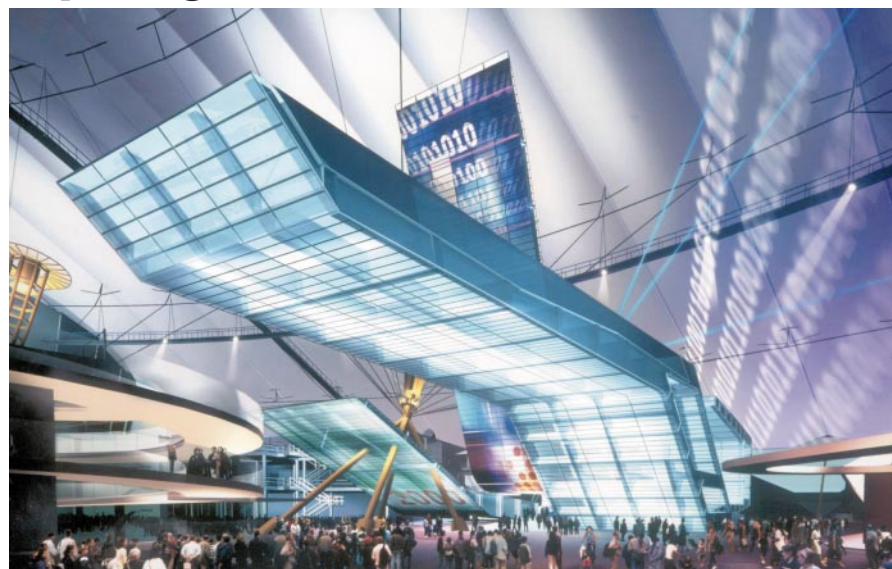
"It is doubtful whether our technology would be capable of achieving the projected target. We feel this should be discussed with the United States," says Taro Nakayama, former minister of foreign affairs and chairman of the working group.

The group's report, released last week, says that Japan must build a centralized system for analysing image data from the satellites in order to make swift policy decisions. Japan's lack of such a system means image processing can take up to five days.

The group also warns that Japan lacks data-processing capability, and must therefore improve its technology with US help. But the Japanese government is at pains to avoid speculation that Japan is exploiting military technology.

Asako Saegusa

Exploring the mind in the Millennium Dome



[LONDON] Plans for the science content of the United Kingdom's Millennium Dome were revealed last week. One area, the Mind zone (pictured), will celebrate "the unique creativity of the human brain, by exploring the nature of our senses and perceptions".

Visitors to the Mind zone will see exhibits on brain imaging (including scans of their own brains), artificial intelligence and the scale of time and space. They will also be able to interact with some of the

earliest intelligent robots — clones of robot tortoises developed for the 1951 Festival of Britain. One of 14 exhibition zones within the dome, the Mind zone is being sponsored by GEC and British Aerospace.

Twice as big as the previous world record holder, the dome will be the centrepiece of the government's Millennium celebrations. Located at Greenwich, it is expected to receive up to 12 million visitors marking the year 2000.

Natasha Loder

IMAGE: HAVES DAVIDSON/NMEC

Uncertainty over leukaemia link to nuclear reactor

[MUNICH] German scientists and politicians last week challenged a study commissioned by a local citizens' group into the north German nuclear reactor 'Krümmel' which concluded that radioactive emissions are the likely cause of a series of leukaemia cases.

Since 1989, 21 leukaemia cases have occurred near Krümmel, and three children have died. A new study carried out by Inge Schmitz-Feuerhake, a physicist at the University of Bremen, found Americium 241, a daughter product of plutonium, in dust samples collected near Krümmel.

Schmitz-Feuerhake argues that emissions from the reactor are the only possible cause of the contamination. But experts at the University of Bremen and the National Research Centre for Environment and Health in Munich say that, since Americium 241 is in the fallout from nuclear weapon tests, it can be detected almost everywhere.

Japan will help to combat air pollution in China

[TOKYO] Japan and China are to collaborate in projects to tackle environmental problems in

China. Projects mooted include developing alternative energy use and curbing greenhouse-gas emissions in urban areas, according to a joint communiqué released by the governments during last week's visit to Japan by Jiang Zemin, the Chinese president.

Japan will help to develop Dalian, Guiyang and Chongqing into "models cities" by applying anti-pollution technologies, setting up programmes to improve energy efficiency and supervising training schemes for energy policy-making. China has indicated its intention to participate in the preliminary activities of the Acid Deposition Monitoring Network in East Asia, a programme chaired by Japan's Environment Agency.

Donation funds German computing centre

[MUNICH] A new institute for computer sciences is to be established next year in Potsdam, east Germany, with the help of a personal donation of DM100 million (US\$59 million) from Hasso Plattner, the head of SAP, Germany's most successful software company. This is the largest private contribution to university education ever made in Germany, where private university funding has little tradition.

The Hasso Plattner Institute will offer bachelor and masters courses in software

systems engineering to 100 advanced informatics students at the University of Potsdam. The state government of Brandenburg, where Potsdam is located, hopes that money from the European Union's 'structural funds' will provide most of the estimated DM40 million construction costs of a new building for the institute.

Moscow plans to boost jobs in science towns

[MOSCOW] The authorities of Moscow district — a vast territory around the Russian capital which excludes Moscow itself — have adopted a programme to support the 24 'scientific towns' in the territory. The scientific organizations which are at the heart of the towns belong to the state, but the federal budget fails to finance them fully.

"The level of unemployment in our scientific towns is two to three times the average for the district," says Alexei Vorontsov, deputy head of the Moscow district administration.

UK systematics report highlights priorities

[LONDON] The first complete assessment of the United Kingdom's natural history collections and expertise in systematics and

biodiversity was launched this week by the UK Systematics Forum.

The report, *The Web of Life*, seeks to focus research on three areas of high priority: description and documentation of biological diversity, the development of classification systems, and the organization of results. Launching the report, the Natural History Museum in London described the British collection of more than 100 million specimens as "a resource unrivalled in its scientific and historical importance".

Societies look at effects of Scottish devolution

[LONDON] The Royal Society and the Royal Society of Edinburgh are to conduct a joint study of the implications for science of the new Scottish parliament. A working group of the two societies, chaired by Geoffrey Boulton, provost and dean of science and engineering at the University of Edinburgh, will issue a report just after the first elections to the parliament are held on 6 May.

The parliament will devise a committee structure and, scientists hope, a means of obtaining appropriate scientific and technical advice. Although it will not have jurisdiction over the UK research councils, which fund most scientific research in Scotland, it will control general university

funding and various research-reliant government functions, including agriculture, industry and fisheries.

NASA satellite to study how a star is born

[WASHINGTON] A two-year NASA mission to study star formation in interstellar clouds was due to get under way this week. Launched by a Pegasus rocket into orbit 600 kilometres above the Earth, the Submillimeter Wave Astronomy Satellite (SWAS) has been planned to observe the spectra of hundreds of interstellar clouds containing water, molecular oxygen, atomic carbon, and carbon monoxide.

Observations in the satellite's sensitivity range from 487 to 556 gigahertz are not possible from the ground because of interference from the Earth's atmosphere.

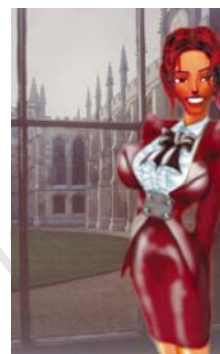
Brussels names advisers for Framework projects

[MUNICH] The European Commission last week appointed 27 experts as members of 17 independent advisory groups, to provide advice on the content and direction of research financed under the fifth Framework programme of research.

The members were selected from among

5,000 proposed candidates. More than one quarter of the selected experts are women, although only ten per cent of the candidates were female. Most members of the advisory groups are academic researchers, and one third come from industry.

Computer game star is science 'ambassador'



[LONDON] Lara Croft (left), the larger-than-life star of the UK-designed computer game "Tomb Raider", has won surprising endorsement from the science minister, Lord Sainsbury, as an "ambassador" for British scientific excellence.

Speaking this week on science and the knowledge economy, Sainsbury argued that Britain should be promoting its contemporary high-tech achievements as much as those of the past. "I want people, when they think of this country, to think of such scientific achievements as Thrust, the first supersonic car, rather than Stephenson or Faraday," Sainsbury said.

Whistle-blower's charter on the way

Sir— The gagging of research scientists against the public interest, of which John Ziman writes, will become more difficult next year, at least in the United Kingdom (*Nature* **395**, 856–857; 1998). The Public Interest Disclosure Act was passed with all-party support in July 1998 and, according to the Department of Trade and Industry, will take effect early next year.

The act protects an employee who discloses that their employer is party to a criminal offence, is failing to fulfil a legal obligation, is endangering health and safety, or is damaging the environment. The disclosure must be made in good faith and not for personal gain. If these conditions are

met, the act declares void any clause in the employee's contract that forbids disclosure. The employee is protected against dismissal for making the disclosure and against "being subjected to any detriment" by the employer. Industrial tribunals will process claims and compensation is currently limited to £12,000 (US\$19,900), but the government is committed to removing this limit.

The employee must take the matter up with the employer first and, if it is not resolved, can then disclose it more widely. Government employees can inform the relevant minister directly. The delay in implementing the act has resulted from the need to make regulations specifying how

employees in other professions should inform and how long they should give employers to act before going public.

The act will create unprecedented transparency in scientific work, as in all other fields of employment. If it had already been in force, for example, the risks of bovine spongiform encephalopathy would have been made public earlier and acted on more quickly, and the mortality rates of patients of certain surgeons at Bristol Royal Infirmary might well have been reported and investigated earlier.

Laurie Smith

23 Pine Walk, Carshalton,
Surrey SM5 4ES, UK

It will take more than notebooks to stop fraud

Sir— In response to your editorial, "Surviving misconduct is one thing, accountability is another", as individuals who have been embroiled in the process we find the recommendations offered to be laudable (*Nature* **395**, 727; 1998). However, we would like to offer the following observations regarding the daunting task of interpreting laboratory records, some dated more than 10 years ago.

It is not sufficient that institutions should require "good laboratory notebooks" to be kept, and archive them for 10 years. More attention should be given to the process by which laboratory data are recorded. Each laboratory and each scientist has different standards and methods of recording data. Notebooks are highly individualized. As a result, even review of the data contained in "good" notebooks after ten years is opaque to those not involved in the experiments, and in many cases to the experimenters themselves.

Even the best notebooks suffer too many lapses in the recording of data to provide an unequivocal picture of the experiments and the results obtained. Such shortcomings make trying to reconstruct what may have gone wrong a decade earlier difficult, if not impossible.

We suggest the adoption of a standardized recording protocol like that used in medicine. Medical records serve to establish standard procedures. Laboratory research records should adhere to guidelines established by the institution for the recording of data obtained from specific procedures. These records should address certain components in the recording of data

from standard procedures used by most labs. In this manner, a uniformity in the recording of data would be established that would serve (as a medical chart does) not only to allow those procedures to be reproduced but also as a detailed record which adheres to a commonly accepted standard for certain lab procedures.

In the absence of standardized guidelines, the review of individualized notebooks and the ability of researchers to adhere to "institutional" targets will fall short of providing a mechanism for preventing misconduct and ensuring accountability in research. Enforcement from above is one thing, but it is equally important to establish consistent procedures for the education of staff in recording data.

This would aid the scientific community and help minimize the adversarial episodes that can emerge from the lack of common guidelines.

Kimón Angelides

Durham University,
Durham DH1 3HP, UK

James V. Pianelli

McGehee and Pianelli, LLP,
1225 N. Loop West, Suite 810,
Houston, Texas 77008-1761, USA

Time to bury misleading myth about careers

Sir— It was with great sadness that I read of the suicide of Jason Altom, a troubled chemistry graduate student caught in the grip of an over-competitive research arena with too little academic guidance (*Nature* **395**, 823 & 826; 1998). The minds of many of our brightest young scientists are still clouded with the notion that, once they

enter the PhD pipeline, there is no escape save for success (tenure-track research in academia) or failure (at best, industrial postgraduate research; at worst, a terminal masters degree).

Too often, the elite academic research system perpetuates this myth through an ignorance of the notion of alternative scientific careers.

I received less immediate support than I expected during my graduate studies, and was forced to seek advice beyond my thesis adviser. Through support from others, I sought an alternative career. I now have a fulfilling career in scientific publishing, where I have been able to combine my love of science and communications.

Similar alternative career possibilities exist in education, technology transfer and law, for those students who are willing to become informed and risk the (undeserved) stigma of leaving the pipeline from another exit (see *Careers and Recruitment* in this issue, pages 493–496).

Your editorial and report outlined some efforts by concerned scientists and academicians to address this issue. Another recent article proposes that it is the responsibility of graduate programmes to provide their students with a range of mentoring opportunities beyond that of the traditional principal investigator and to ensure that they have access to support services, but students must also take more responsibility for their own support needs (*New Anat.* **253**, 132–134; 1998).

Surely, much has been written on this subject since the PhD glut of the 1990s took centre stage among the worries of our next generation of researchers. Now, words are no longer enough; action is essential.

Mark H. Paalman

The Anatomical Record,
3241 Blackwalnut Drive,
Annapolis, Maryland 21403-4652, USA

An oblique view of climate

Bruce G. Bills

One explanation for certain patterns of glaciation in the past invokes a large and comparatively swift decline in the tilt, or obliquity, of the Earth. A provocative hypothesis provides a mechanism by which such a decline could have occurred.

How do variations in Earth's orbital and rotational geometry influence climate? Does the climate system, in turn, influence rotation? We all experience the radiative and thermal cycles of night and day, winter and summer. So we are familiar enough with the influence of Earth's rotational and orbital motions on the spatio-temporal pattern of light and temperature to make it easy to imagine how long-term variations in the orbit and rotation would affect climate. Much recent data and modelling help confirm that principle^{1,2}.

Somewhat further removed from human experience is the notion that climatic change itself could influence the rotational dynamics of the Earth. That, however, is the hypothesis advanced by Williams *et al.* on page 453 of this issue³. The basic idea is quite simple, and involves feedback between Earth's obliquity (the angular separation between the spin pole and orbit pole around the Sun, Fig. 1) and its oblateness (departure from spherical symmetry).

First, however, it is useful to recall how orbital and rotational geometry influences climate. The main seasonal cycle is primarily determined by the orientation of the spin

axis, and only secondarily by the eccentricity of the orbit. Currently, perihelion (Earth's closest approach to the Sun) occurs several weeks after winter solstice in the Northern Hemisphere (shortest daylight). However, neither the orbit nor the spin axis is fixed in space. Gravitational interaction with other planets (principally Venus) causes the shape and orientation of the orbit to change on a variety of timescales⁴, with the dominant period near 70,000 years (70 kyr) and subsidiary oscillations at periods ranging from 50 kyr to 1.9 million years (Myr).

Gravitational torques exerted by the Moon and Sun on the oblate figure of the Earth cause the spin axis to precess with a period of 25.8 kyr. If the orbit plane were fixed, the path of the spin pole would be a circle centred on the orbit pole, keeping the obliquity fixed. However, because the orbit is also precessing, the obliquity oscillates by $\pm 1^\circ$ about its present value of 23.5° , with a period of 41 kyr. These obliquity oscillations modulate the seasonal and latitudinal pattern of incident radiation, and thus affect climatic variations.

How then does climate change influence rotation? One way is to change the spin

precession rate by changing the oblateness of the Earth's mass distribution. During major glacial cycles, mass transport between the oceans and ice sheets is sufficient to change the precession rate by about 1%. The net change includes accumulation of continental ice and partially compensating subsidence of the Earth's surface. If the obliquity and oblateness oscillations are exactly in phase, there is no long-term net effect. But if the oblateness lags behind the obliquity, there will be a secular change in obliquity, with a rate that depends on the amplitude and phase of the oblateness variations⁵⁻⁷.

The long-term stability of Earth's climate system is an important question, but one that remains elusive. Despite progress in short-term weather prediction (based on improved quality and quantity of observations, faster computers and better understanding of the system dynamics), our understanding of long-term climate dynamics is still quite primitive. Part of the problem, of course, is that the further back into the past we go, the more difficult it is to reconstruct which path the climate system has followed. When we still don't know what has happened, how can we reconstruct why? In this situation, the role of theoretical models such as that of Williams *et al.*³ is not so much to explain what actually happened as to broaden our perspectives on the types of behaviours that might have occurred.

The principal novelty of the new work by Williams *et al.* concerns the rate and direction of the obliquity change they seek to explain (Fig. 1). They are invoking an obliquity-oblateness feedback to decrease Earth's obliquity, rather than increase it, and the obliquity change they seek is $20-30^\circ$ in less than 100 Myr. They propose that such a change could have been caused by the very large ice sheets that might have built up during periods when the continents were clustered near to the poles. The motivation for that particular scheme comes from a suggestion that Earth's obliquity has been large throughout much of its history⁸ (in excess of 54° , the value at which the annual average incident radiation at the Equator equals that at the poles; at greater obliquities, the equatorial regions become colder than the poles). That suggestion was made in an attempt to explain evidence for low-latitude glacial deposits from the Neoproterozoic (750–550 Myr ago), while avoiding a 'snowball Earth', in which the entire surface of the Earth would be covered by ice⁹⁻¹².

As always, more work is needed. In this case, distinguishing between the two competing climatic possibilities (equatorial versus global glaciation) should be easily resolvable by searching for contemporaneous high-latitude and low-latitude glacial deposits¹³. Reconstructing an unambiguous obliquity history will be more of a challenge. □

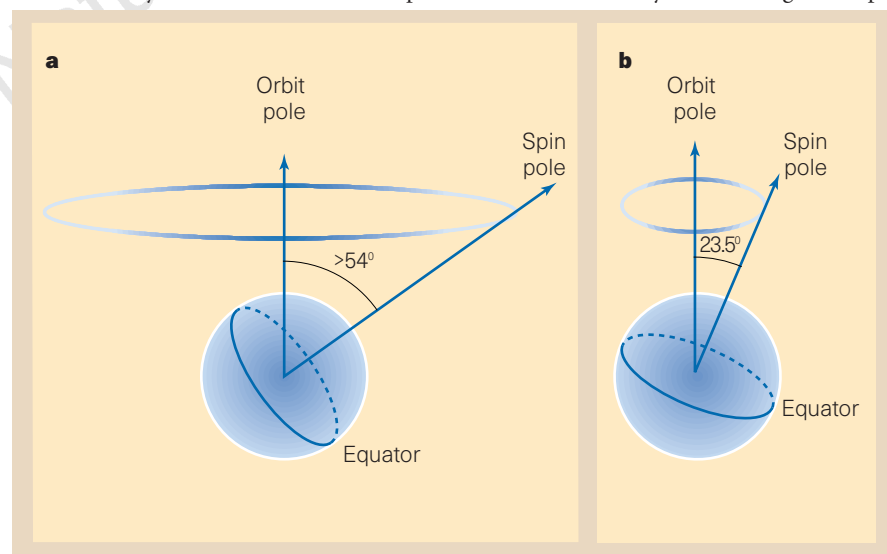


Figure 1 Earth's obliquity. Williams and colleagues' work³ stems from the need to account for a $20-30^\circ$ decline in obliquity, required by a hypothesis that explains the occurrence of low-latitude glacial deposits 750–550 Myr ago through an obliquity of more than 54° (a), when the equatorial regions would have been colder than the poles. b, Since about 430 Myr, obliquity is thought to have been around 23.5° . The authors propose that ice-sheet formation and configuration (not shown) was the driving force.

Bruce G. Bills, currently at the Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, is in the Geodynamics Branch, NASA Goddard Space Flight Center, Mailstop 921.0, Greenbelt, Maryland 20771, USA. e-mail: bills@denali.gsfc.nasa.gov

1. Hilgen, F. J. *Earth Planet. Sci. Lett.* **107**, 349–368 (1991).
2. Imbrie, J. *et al.* *Paleoceanography* **7**, 701–738 (1992).
3. Williams, D. M., Kasting, J. F. & Frakes, L. A. *Nature* **396**, 453–455 (1998).
4. Laskar, J. *Astron. Astrophys.* **198**, 341–362 (1986).

5. Rubincam, D. P. *J. Geophys. Res.* **98**, 10827–10832 (1993).
6. Bills, B. G. *Geophys. Res. Lett.* **21**, 177–180 (1994).
7. Ito, T., Masuda, K., Hamano, Y. & Matsui, T. *J. Geophys. Res.* **100**, 15147–15161 (1995).
8. Williams, G. E. *Earth Sci. Rev.* **34**, 1–45 (1993).
9. Harland, W. B. *Geol. Rundsch.* **54**, 45–55 (1964).
10. Caldeira, K. & Kasting, J. F. *Nature* **359**, 226–228 (1992).
11. Evans, D. A., Beukes, N. J. & Kirschvink, J. L. *Nature* **386**, 262–266 (1997).
12. Hoffman, P. F., Kaufman, A. J., Halverson, G. P. & Schrag, D. P. *Science* **281**, 1342–1346 (1998).
13. Kennedy, M. J. *et al.* *Geology* **26**, 1059–1063 (1998).

Developmental biology

Attractive genetics

Ken Howard

On page 466 of this issue, Van Doren *et al.*¹ report the identification and analysis of a gene in *Drosophila* embryos that attracts the precursors for sperm and eggs — the germ cells — as they migrate to the gonad in order to mature. Christened *columbus*, this is the first gene to be implicated in producing an attractant for germ cells in this model system. The authors show that *columbus* encodes 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase², an enzyme best known for its role in cholesterol biosynthesis in humans³, and the target of cholesterol-lowering drugs. HMG-CoA reductase is unlikely to be the attractant itself because it is located intracellularly, in the membrane of the endoplasmic reticulum, but it defines a new pathway for producing an attractant that may well be a lipid.

Earlier this year, the Lehmann lab reported an exhaustive genetic screen of *Drosophila* for functions required during germ-cell migration and gonad formation⁴, and *columbus* is one of the handful of genes that they identified. Several of these genes encode nuclear factors that regulate genes involved in the formation of gonadal target tissue⁵.

But mutations that disturb germ-cell migration without preventing formation of the target tissue should point the way to pathways involved in the production of specific signals to guide the germ cells.

This is just what *columbus* does. Van Doren *et al.* found that in mutant embryos where it is not expressed, germ cells no longer move towards the perfectly formed gonad. The simplest model would be that HMG-CoA reductase produces a factor that attracts germ cells — a proposal that is supported by the fact that the RNA for this enzyme is expressed at high levels in the gonad. To test this hypothesis the authors expressed HMG-CoA reductase at high levels in other tissues, and found this made tissues that never normally associate with germ cells attractive to them. For instance, if a nerve cord is genetically reprogrammed to express high levels of HMG-CoA reductase, it attracts many germ cells that choose the nerve cord, instead of the gonad, as a target (Fig. 1).

The immediate biochemical product of the reaction catalysed by HMG-CoA reductase is mevalonate, which is used in humans to produce cholesterol. However, flies don't

use mevalonate to make cholesterol — instead, it feeds into the production of non-sterol isoprenoids, including bioactive juvenile hormones. Although the pathway from HMG-CoA reductase to the germ-cell attractant is not known, perhaps a lipid produced from mevalonate will turn out to be that attractant.

There are a number of cases where lipids and lipid-modifying enzymes are involved in controlling cell migration. Genetically, one of the best characterized is the role of the lipid platelet activating factor (PAF) in neuronal migration during development of the human brain. This function was discovered when a gene involved in a human genetic disease, Miller–Dieker lissencephaly, was found⁶ to encode the α -subunit of an enzyme that inactivates PAF, PAF acetyl hydrolase. Other work in this system points to a simple mechanism whereby PAF modulates migration negatively, and the acetyl hydrolase generates low points in a gradient of PAF that guides migrating neurons⁷. Another example of a lipid-modifying enzyme that controls cell migration is a *Drosophila* lipid phosphatase⁸, which is involved in repelling the same germ cells that high levels of HMG-CoA reductase attract.

Although the active product of the *columbus* pathway may be a lipid, it is equally possible that high levels of HMG-CoA reductase are involved in a protein-modification reaction that is required for biological activity. Protein modification has already been associated with signalling in flies. For example, a putative glycosyltransferase encoded by the *fringe* gene is involved in Notch-dependent signalling⁹, and the glycosaminoglycan-modifying enzyme heparan sulphate 2-O-sulphotransferase encoded by the *pipe* gene is involved in dorso-ventral patterning¹⁰.

What is the final product of the *columbus* pathway? And how does this molecule attract germ cells? The attractant could be anything from a specifically modified protein to a lipid. If it turns out to be a hitherto unknown molecule, it will undoubtedly be investigated for bioactivity in other, perhaps more clinically relevant, contexts. If, on the other hand, the molecule is a familiar one showing a previously unsuspected activity, we will have to re-evaluate the molecule's biological functions. Either way, the connection between HMG-CoA reductase and a substance that attracts germ cells may point to new therapeutic agents or new uses for existing ones. It will certainly lead to exciting and challenging experiments to define the attractive product that this enzyme makes in order to draw germ cells to the gonad. □

Ken Howard is at the MRC Laboratory for Molecular Cell Biology and the Department of Physiology, University College London, Gower Street, London WC1E 6BT, UK.

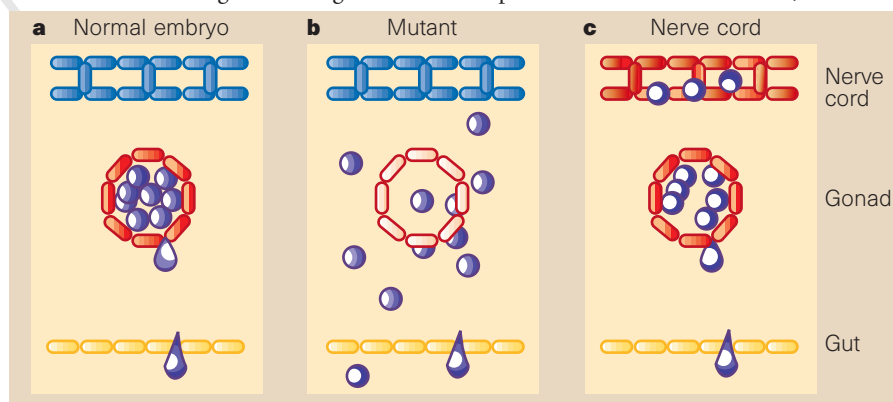


Figure 1 Action of HMG-CoA reductase on germ-cell migration. a, Normal migration of the germ cells when they move from inside the gut (yellow cells) to the gonad. Van Doren *et al.*¹ have found that the gonad expresses high levels of HMG-CoA reductase (red). The germ cells ignore the nerve cord, which does not express this enzyme at high levels. b, In an HMG-CoA reductase mutant, germ cells still move through the gut into the body cavity but few of them end up in the gonad. c, In an experimental situation where HMG-CoA reductase is expressed at high levels in the nerve cord, the germ cells will home there too, attracted by a product of the HMG-CoA enzyme activity.

e-mail: ken.howard@ucl.ac.uk

1. Van Doren, M., Broihier, H. T., Moore, L. A. & Lehmann, R. *Nature* **396**, 466–469 (1998).
2. Gertler, F. B., Chiu, C. Y., Richter-Mann, L. & Chin, D. J. *Mol. Cell. Biol.* **8**, 2713–2721 (1988).
3. Goldstein, J. L. & Brown, M. S. *Nature* **343**, 425–430 (1990).
4. Moore, L. A., Broihier, H. T., Van Doren, M., Lunsford, L. B. & Lehmann, R. *Development* **125**, 667–678 (1998).
5. Broihier, H. T., Moore, L. A., Van Doren, M., Newman, S. &

- Lehmann, R. *Development* **125**, 655–666 (1998).
6. Hattori, M., Adachi, H., Tsujimoto, M., Arai, H. & Inoue, K. *Nature* **370**, 216–218 (1994).
7. Albrecht, U. *et al. Dev. Biol.* **180**, 579–593 (1996).
8. Zhang, N., Zhang, J., Purcell, K. J., Cheng, Y. & Howard, K. *Nature* **385**, 64–67 (1997).
9. Cohen, B. *et al. Nature Genet.* **16**, 283–288 (1997).
10. Sen, J., Goltz, J. S., Stevens, L. & Stein, D. *Cell* **95**, 471–481 (1998).

Particle physics

Time's broken arrow

Ken Peach

Two experiments^{1–4} studying the properties of a particle called the neutral kaon have reported clear evidence that the world is not symmetric with respect to time—that is, it does matter which direction the arrow of time chooses to take. These observations are not unexpected, but they show that the quantum world still has mysteries to be revealed.

Although special relativity teaches us that time and space are intimately entwined, time is still a very special variable. For example, in space we can compare the results of an experiment that is aligned north–south in one place with those of the same experiment aligned south–north or east–west in another. The classic Michelson–Morley experiment, which showed that the velocity of light does not depend upon the direction of space, and which was so crucial in the development of special relativity, is just one example of an experiment which relies on measurements in different regions of and directions in space. Time, however, is different. We can compare the results of an experiment today with those obtained yesterday, but we cannot so easily compare the results of experiments with different directions of time.

The two experiments manage to circumvent this by using the properties of the neutral kaon. The neutral kaon is made from a down quark and an antistrange quark, whereas the neutral antikaon is made from an antidown quark and a strange quark. They are different particles. However, once created, some quantum-mechanical ‘magic’ occurs so that as time passes an initially pure

kaon state becomes a mixture of kaon and antikaon states, a process known as mixing. Similarly, an initially pure antikaon state will also become mixed.

In a universe in which it is immaterial which is called the kaon (matter) and which is called the antikaon (antimatter), a beam of neutral kaons would evolve into a mixed state which consisted of exactly equal amounts of kaon and antikaon, independently of the production mechanism. Such a universe would also be symmetric under the reversal of time.

It has been known for more than 30 years that the properties of matter and antimatter are not completely equivalent. In particular, in any beam of neutral kaons there will eventually be a small (0.3%) excess of matter over antimatter. This is the phenomenon of CP-violation, where C is the quantum-mechanical operation of replacing all particles by their antiparticles and P is the parity operation of reversing the direction of all spatial components. There is strong evidence to support the theoretical expectation that the Universe should be symmetric if the CP operation is combined with T, the time-reversal operation which reverses the direction of time. (There is a subtlety in that this operation also changes the quantum-mechanical amplitude which describes the physical process into its complex conjugate.) Until now, there has been no direct evidence for an asymmetry under time reversal.

The CPLEAR experiment at CERN, Geneva, has reported convincing evidence for such an asymmetry^{1,2}. The concept is

quite simple — produce kaons and antikaons and observe them, recording when they decay as antikaons and kaons respectively (Fig. 1). If the probability that an initially produced kaon decaying as an antikaon is different from that of an initially produced antikaon decaying as a kaon, then self-evidently the process is not symmetrical under time reversal.

The experiment relies on being able to ‘tag’ the matter/antimatter state of both the produced kaon and the decayed kaon. The kaons and antikaons are generated in proton–antiproton annihilations, where the initial state is identified by noting the strangeness of the accompanying charged kaon. A neutral kaon is accompanied by a negatively charged kaon while the neutral antikaon is accompanied by a positively charged kaon.

When the neutral kaon decays into a pion, an electron and a neutrino, the electron charge is positive (that is, e^+ , a positron), whereas the antikaon decays to a negatively charged electron, e^- .

The CPLEAR team has observed a total of 1.3 million events in which a kaon demonstrably changes its nature. It turns out that there is a $0.66 \pm 0.18\%$ greater probability that an antikaon will turn into a kaon than the other way round.

Meanwhile, as reported at a workshop and described on a website^{3,4}, the KTeV experiment at Fermilab, Chicago, has come to a similar conclusion using a very different method. By studying the dynamics of the decay of the neutral kaons into two charged pions and an electron–positron pair, it is possible to deduce time-reversal non-invariance. In effect, the distribution of the angle between the plane containing the electron–positron pair and the plane containing the positive and negative pion pair should be symmetrical if time reversal is a good symmetry. From a data sample of 1,800 such decays, identified in a total of 130 million events, they observe an asymmetry of $13.5 \pm 3.9\%$, clear evidence that time reversal is also violated in this decay.

The results of both experiments are consistent with expectations based on the observed amount of CP violation in the neutral kaon system. Further measurements will be possible, for example at the experiments that begin next year at the Stanford Linear Accelerator, California, and at KEK, the accelerator facility at Tsukuba in Japan. These will follow decay of the B meson, which is similar to the kaon but with the strange quark replaced by a bottom quark. This system should have similar characteristics to the kaon, and should show similar effects. There is also the possibility that the measurements of the electric dipole moment of the neutron could reveal evidence for time-reversal non-invariance.

What is exciting now is that, for the first

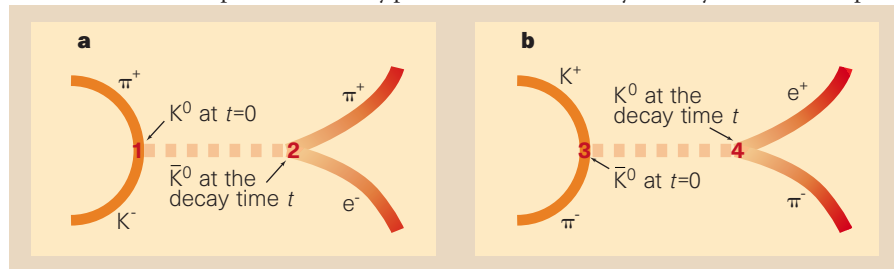


Figure 1 Principle of the CPLEAR measurement. a, At point 1, proton–antiproton annihilation produces a positively charged pion (π^+), a charged kaon (K^-) and a neutral kaon (K^0). But when K^0 decays at point 2 at time t later it does so through a decay characteristic of the antikaon ($\bar{K}^0$). b, An antikaon is produced in a similar way at point 3, and when it decays at point 4, time t later, it does so through characteristic kaon decay. a and b are time reversed processes; the greater probability of the latter shows that the direction of time's arrow does indeed matter.

time, there is direct evidence that the direction of time's arrow does matter — the Universe would be different if the universal clock were the other way round. □

Ken Peach is in the Department of Particle Physics, CLRC-Rutherford Appleton Laboratory, Chilton, Nr

Didcot, Oxfordshire OX11 0QX, UK.

e-mail: Ken.Peach@rl.ac.uk

1. <http://www.cern.ch/cplear/Welcome.html>
2. Angelopoulos, A. *et al.* *Phys. Lett. B* (in the press).
3. Arenton, M. *Workshop on Heavy Quarks at Fixed Targets*, Fermi National Accelerator Laboratory, Chicago, 10–12 October 1998.
4. http://fnphyx-www.fnal.gov/experiments/ktev/pipice/pipice_bg.html

Population biology

The voles of Hokkaido

Robert M. May

Lemmings do not jump off cliffs into the sea in episodes of cult-like mass suicide, unless herded to do so by Walt Disney's cameramen (as allegedly happened in one of his early natural history films). In some places, however, their abundance does exhibit roughly 4-year cycles, with peak years seeing as much as thousandfold increases over low years; in other places, the population fluctuations are simply seasonal.

Elton^{1,2} first drew attention to such cycles in the abundances of mice, voles and lemmings in northern regions, asking what they can tell us about the dynamics of, and the underlying regulatory mechanisms in, natural populations. In the middle years of the century much work was done on the natural history of such cycles, particularly in Fennoscandia. Largely unknown in the west, possibly the best such data set has quietly been accumulating since the 1920s on the vole *Clethrionomys rufocanus* (pictured) of Hokkaido, the northernmost island of Japan. The 'discovery' of these data by the Norwegian ecologist Nils Stenseth and his colleagues has catapulted the data to centre stage, and the subsequent work is the subject of a collection of papers in a special feature in the Japanese journal *Researches on Population Ecology*³.

As one contributor (C. J. Krebs) points out, there are three reasons why such rodents are ideal subjects for studying population dynamics. First, they are found virtually everywhere, and are often pests in forest plantations. Second, their short lives mean significant work can be done on the 3–4-year timescale of the PhD. Third, as Elton emphasized, they do interesting things, like cycling.

Initial studies of the Hokkaido voles were carried out by Kinoshita⁴, and indeed arose from the problems they caused foresters. This led Ota and Ueda to develop a systematic monitoring programme, beginning in 1964, which has generated "the world's most extensive and standardized set of time series data on small rodents covering 225 locations over ... Hokkaido for a period of up to 31 years (with recordings made three times a year; spring, summer and fall)" (N. C. Stenseth and I. Saitoh). It is a sad irony that this data stream faded out in 1995, owing to lack of interest and funds, just as Stenseth and others came on the scene. One aim of the

special feature is to see the programme reinstated.

The data, when properly analysed, can show us patterns in the population dynamics. But the real ecological interest is thence to elucidate the underlying processes — the interactions among individuals and between species — which generate such patterns. Advances in understanding the dynamical behaviour of nonlinear systems have, over the past 25 years or so, helped us understand why populations of animals may variously exhibit roughly steady abundance, or regular cycles, or unpredictable fluctuations driven by internal mechanisms ('chaos'), external environmental noise or a mixture of the two^{5–7}. In particular, methods have been developed for analysing fluctuating time-series, with the aim of asking how many interacting components are likely to be needed to explain the observed patterns in a population's abundance. In other words, and very roughly, we seek to infer from the data how many distinct regulatory factors — how many independent variables; how many other populations — are relevant^{8–10}.

For example, a notable application¹¹ of these methods to the more than 100-year-old data stream on the 11-year cycles in Canadian lynx and snowshoe hare suggests one external factor for the lynx dynamics (the hares?) and two for the hare dynamics (lynx and food?). The problem with all such analyses is that, although they advance beyond mere patterns, to suggest the number of independent factors associated with regulation of population density, they give no direct hint as to what these biological factors are. This, incidentally, sheds light on a long-standing controversy in population biology: is the lynx–hare cycle driven by the predator–prey interaction, or by an interaction between hares and their food (or some equivalent factor, such as parasites)? Stenseth's analysis bears out earlier suggestions⁵ that this is a false dichotomy, with both predators and food being entwined in the explanation.

The above methods need long time-series if they are to give reliable results. The 31-year sweep of

the longest Hokkaido time-series makes them short (less than ten times the length of the 3–5-year vole cycles), but offsetting this is the large number of more-or-less independent series for different locations, 194 of which have a length of 23 years or more. At individual locations, population fluctuations ranged from as small as around tenfold to as large as around one-thousandfold. Analysis of these data (T. Saitoh *et al.*) shows evidence both for direct density-dependent mechanisms and for mechanisms operating with one-year delays (corresponding crudely to one other variable, or population interaction, being involved in vole regulation). Both direct and delayed density-dependence tended to be strongest in more northern and eastern populations.

The fact that the Hokkaido data come from 225 different places makes them especially interesting. Broadly, sites in the north and east (where, as seen above, density dependence seems most marked) tend to show pronounced cycles, with average periods in the range 3.5–4.5 years. Sites more to the south and west showed relative stability in average populations from year to year (annual cycles). An analysis of 185 of the sites found the typical spatial scale over which fluctuations are synchronized between sites, as measured by the distance at which similarity is equal to that expected by chance alone in the region, to be around 50 km (O. N. Bjørnstad *et al.*). The authors see this as a puzzle, because of their estimate that a local population does not extend over distances greater than about 8 km, based on inter-generational dispersal. I see no such puzzle, because I think this analysis is too simplistic; very tiny amounts of spatial movement between cities can, in both theory and practice, lock cycles in the incidence of measles and other childhood infections into rough synchrony (ref. 12; B. T. Grenfell and B. F. Finkenstädt).

The special feature³ goes beyond temporal and spatial dynamics, to discuss the taxonomic status of the Hokkaido vole (Y. Kaneko *et al.*); its social organization (Y. Ishibashi *et al.*); and mark–recapture experiments, where there are pointers to further work needed if we are to understand the detailed mechanisms that regulate vole abundance in Hokkaido, and elsewhere (N. G. Yoccoz *et al.*). The volume concludes with

brief essays from the Great and Good of population ecology, the most interesting of which draw illuminating parallels with quasicyclic dynamics in other areas: in the incidence of childhood infections before vaccination, and in Soay sheep numbers (Grenfell and Finkenstädt), and in



insect populations (M. Begon).

Stenseth's papers in the collection are peppered with epigrammatic quotations. My favourite is "Mathematics without natural history is sterile, but natural history without mathematics is muddled"¹³. More crudely, "Sound naturalism is to ecology what legs are to a runner; but anti-theoretical naturalists are, quite naturally, like headless runners"¹⁴. This collection walks tall, from head to toe. □

Robert M. May is in the Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

e-mail: robert.may@zoo.ox.ac.uk

1. Elton, C. S. *J. Exp. Biol.* **2**, 119–163 (1924).

2. Elton, C. S. *Voles, Mice and Lemmings* (Clarendon, Oxford, 1942).
3. Stenseth, N. C. & Saitoh, T. (eds) *Res. Pop. Ecol.* **40**, 1–158 (1998).
4. Kinoshita, E. *Res. Bull. College Expt. Forests, Hokkaido University* (Sapporo) **5**, 1–5 (1928) (in Japanese).
5. May, R. M. *Stability and Complexity in Model Ecosystems* (Princeton Univ. Press, 1973).
6. May, R. M. *Nature* **261**, 459–467 (1976).
7. Turchin, P. in *Population Dynamics* (eds Cappuccino, N. & Price, P.) 14–40 (Academic, New York, 1995).
8. Sugihara, G. & May, R. M. *Nature* **344**, 734–741 (1990).
9. Sugihara, G. *Phil. Trans. R. Soc. Lond. A* **348**, 477–495 (1994).
10. Tong, H. *Scand. J. Stats* **22**, 399–421 (1995).
11. Stenseth, N. C. *et al. Proc. Natl Acad. Sci. USA* **94**, 5147–5152 (1997).
12. Lloyd, A. L. & May, R. M. *J. Theor. Biol.* **179**, 1–11 (1996).
13. Maynard Smith, J. *New York Review of Books* 41–42, 13 May (1983).
14. Haila, Y. & Jarvinen, O. in *Conceptual Issues in Ecology* (ed. Saarinen, E.) 261–278 (Reidel, Dordrecht, 1982).

Transcription

A lesson in sharing?

Patrick A. Grant and Jerry L. Workman

The TATA-binding protein (TBP) is pivotal to transcription — it recognizes a specific DNA sequence, the TATA element, found in the control region of many genes, and is required for other proteins to assemble there and initiate transcription. A topic of much discussion has been to what degree the TAFs (TBP-associated factors) are required for TBP to do this. Initially TAFs were thought to be essential, but this idea was challenged by the observation that,

in yeast, some TAFs are not generally required for transcription^{1,2}. Moreover, a subset of the TAFs also contribute to the activities of other transcriptional regulatory complexes^{3–6}. Now, four papers in *Molecular Cell*^{7–10} and one report in *Cell*¹¹ fuel this debate by showing that a sub-group of TAF proteins, histone-like TAFs, may have a much broader role in gene expression.

Considerable work over the past decade led to the conclusion that TFIID — an evolu-

tionarily conserved complex consisting of TBP and eight to twelve associated TAFs — rather than TBP alone is required for activation of protein-coding genes. Consistent with this, almost all TAFs were found to be essential for yeast viability. However, this model seemed all but destroyed when experiments to inactivate or deplete certain TAFs in yeast cells did not lead to the expected broad loss of activated transcription^{1,2}. Moreover, TAF145/130, which was thought to be central to the assembly of TFIID, was found to regulate only a subset of genes, further deflating enthusiasm for a model of general TAF function¹².

Now, in one of the new papers, Holstege *et al.*¹¹ describe a transcriptional analysis of most of the yeast genome. They find that TAF145 is required for transcription of about 16% of the genes studied. Surprisingly, parallel analyses of another TFIID component, TAF17 (a histone-like TAF), by Apone *et al.*⁷, Michel *et al.*⁸ and Moqtaderi *et al.*⁹ indicate that it is broadly required, regulating the expression of at least 67% of all yeast genes^{7,11}.

So what distinguishes TAF17 from TAF145? Sequence analysis reveals some similarity between TAF17, TAF60 and TAF68/61, and the histones H3, H4 and H2B, respectively. Perhaps these histone-like TAFs are important structurally, in a complex, through mutual interactions of their histone-fold motifs. Michel *et al.*⁸

Zoology

A saola poses for the camera

As wildlife photographs go, it will not win any prizes for artistry, but as a wildlife self-portrait it is unique. This is the first picture of a wild saola (*Pseudoryx nghetinhensis*, also known as the Vu Quang bovid), a creature which caused a stir five years ago as the first large land vertebrate to be discovered for over 50 years (V. V. Dung *et al. Nature* **363**, 443–445; 1993).

The saola was originally described from skulls, teeth and skins alone, the hunting trophies of local inhabitants collected by a survey team working in the mountainous forests along the Vietnam–Laos border. The form of the adult animal was pieced together from fragments of 20 specimens, from which it was surmised that saolas weigh about 100 kg, and stand one metre tall.

As the artist's reconstruction by Karen Phillipps shows, the saola is an unusual antelope. Its long, straight horns and coloration make it difficult to group with other species, but based on anatomical and DNA evidence it was assigned to a new genus in the bovine group containing the oxen and the eland. Since then, however, a study based on skull form and dentition



has sought to place it with the goats (H. Thomas, *Mammalia* **58**, 453–481; 1994).

Scientific eyes have still to gaze upon a live, wild saola in the flesh. This picture was taken when an obliging animal triggered a camera trap laid by a team led by Mike Baltzer of Flora and Fauna International, which is surveying the forest as part of an EU-funded project to halt its

degradation and destruction. Based on rough calculations of the amount of habitat needed to support an animal of this size, it is thought that probably only a few hundred saola survive in the region. Under attack as it is from hunters and shrinking forests, this odd, elusive creature may well become extinct before any researcher sees one or learns its habits. **John Whitfield**

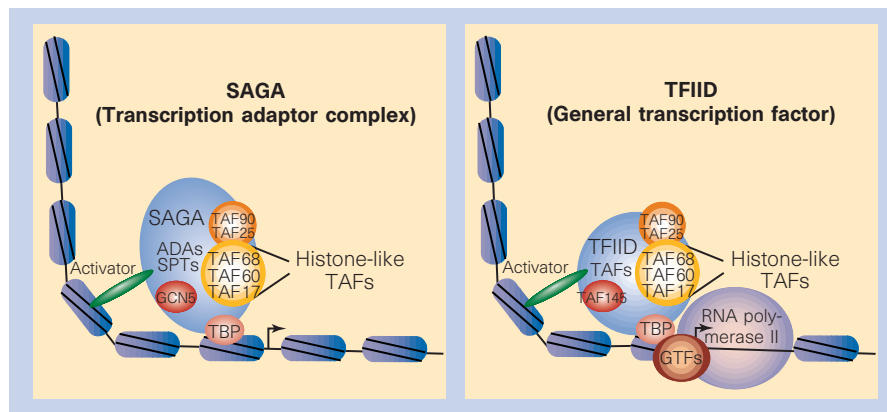


Figure 1 Transcriptional complexes bound to an array of nucleosomes. Both SAGA (SPT–ADA–GCN5 acetyltransferase) and TFIID contain the histone-like TAFs (yellow) and two additional TAFs, TAF25 and TAF90. Each complex contains a subunit with histone acetyltransferase activity (GCN5 in SAGA and TAF145 in TFIID) that is thought to lead to acetylation of histones in promoter proximal nucleosomes (purple). TFIID also contains additional TAFs, whereas SAGA contains ADA and SPT proteins. Both SAGA and TFIID can interact with transcriptional activators, and they bind or contain the TATA-binding protein, TBP. TFIID is also crucial in the formation of transcription complexes, including additional general transcription factors (GTFs) and RNA polymerase II.

examined the importance of histone-like TAFs by inactivating each of these proteins in yeast. These studies, along with inactivation of TAF68 by Natarajan *et al.*¹⁰, point to very general defects in transcription when histone-like TAFs are lost in yeast cells, and provide genetic evidence for interactions between these TAFs⁸. Although previous studies indicate that TAF60 and TAF68 do not have such broad effects¹², the new studies imply that the histone-like TAFs are critical for the expression of most protein-coding genes.

These observations raise a dilemma — why are certain TAFs required for the transcription of more genes than others, if they are all subunits of the same TFIID complex? A possible explanation has come from studies of the yeast SAGA histone acetyltransferase complex, which contains a subset of TAF proteins including all three histone-like TAFs³ (Fig. 1). This arrangement seems to be conserved in homologous human acetyltransferase complexes^{4,6}. SAGA is a large complex of around 20 different proteins, and it interacts directly with gene-specific activators and TBP^{3,10,13–15}. Thus, a logical explanation for the broader transcriptional defect when histone-like TAFs (such as TAF17) are lost, but not when others (such as TAF145) are inactivated, is that the histone-like TAFs occur in other transcriptional regulatory complexes as well as TFIID.

It seems unlikely that the broad transcriptional defect observed when TAF17 is inactivated is simply due to a loss of SAGA activity. For example, GCN5, the catalytic subunit of SAGA, is required for the expression of only about 5% of all yeast genes¹¹. Moreover, inactivation of another component of SAGA does not cause a general decrease in gene expression⁸. However, it is possible that the functions of TFIID and

SAGA in transcription are redundant. So, only when shared components are lost is there a deficiency in both complexes, resulting in broad transcriptional defects. Loss of subunits unique to either TFIID or SAGA would then affect only specific subsets of genes¹¹. This ‘redundancy’ theory is intriguing, because both TAF145 and GCN5 contain histone acetyltransferase activity. However, histone-like TAFs could exist in a third, as-yet-unidentified complex that is distinct from SAGA and TFIID. Or, perhaps the histone-like TAFs have a specific, essential function that is not shared by the other components of SAGA and TFIID. It remains to be seen whether functional redundancy exists, or whether the other, equally plausible, hypotheses hold true¹². □

Patrick A. Grant and Jerry L. Workman are at the Howard Hughes Medical Institute, Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, Pennsylvania 16802-4500, USA.

e-mail: jlw10@psu.edu

1. Moqtaderi, Z., Bai, Y., Poon, D., Weil, P. A. & Struhl, K. *Nature* **383**, 188–191 (1996).
2. Walker, S. S., Reese, J. C., Apone, L. M. & Green, M. R. *Nature* **383**, 185–188 (1996).
3. Grant, P. A. *et al.* *Cell* **94**, 45–53 (1998).
4. Ogryzko, V. V. *et al.* *Cell* **94**, 35–44 (1998).
5. Wicczorek, E., Brand, M., Jacq, X. & Tora, L. *Nature* **393**, 187–191 (1998).
6. Martinez, E., Kundu, T. K., Fu, J. & Roeder, R. G. *J. Biol. Chem.* **273**, 23781–23785 (1998).
7. Apone, L. M. *et al.* *Mol. Cell* **2**, 653–661 (1998).
8. Michel, B., Komarnitsky, P. & Buratowski, S. *Mol. Cell* **2**, 663–673 (1998).
9. Moqtaderi, Z., Keaveney, M. & Struhl, K. *Mol. Cell* **2**, 675–682 (1998).
10. Natarajan, K., Jackson, B. M., Rhee, E. & Hinnebusch, A. G. *Mol. Cell* **2**, 683–692 (1998).
11. Holstege, F. C. P. *et al.* *Cell* **95**, 717–728 (1998).
12. Hahn, S. *Cell* **95**, 579–582 (1998).
13. Saleh, A., Lang, V., Cook, R. & Brandl, C. J. *J. Biol. Chem.* **272**, 5571–5578 (1997).
14. Utley, R. T. *et al.* *Nature* **394**, 498–502 (1998).
15. Sterner, D. E. *et al.* *Mol. Cell. Biol.* (in the press).



100 YEARS AGO

For eight years I have had a large plant of *Atropa* growing here in my garden amongst currants and gooseberries; close by it is a mountain-ash, and at a short distance a large cherry-tree. Birds, including the blackbird, build in the garden; but although the cherries, currants, gooseberries and raspberries are annually stripped, the *Belladonna* berries are never touched. The birds are encouraged, and the fruit can be spared. The *Belladonna* berries are conspicuous objects from July to November; there are hundreds on my plant every year, long after other fruits have vanished — black, lustrous, luscious-looking — but no bird ever touches them.

From *Nature* 1 December 1898.

50 YEARS AGO

One of the paradoxes of civilised life is the lowly position occupied among the sciences by the study of living things. There seems, in Great Britain at least, to be a general disposition to regard biology as something in the nature of a pastime, while the physicist and the chemist are held in respect as serious practitioners. Only in the treatment and, to a modest degree, the prevention of disease, and in agriculture, has biological science a recognized position. That the well-being of the human race is closely bound up with the equilibrium of Nature, is not generally recognized. It was therefore, with special interest that we studied the presidential address of Sir Henry Tizard at the recent meeting in Brighton of the British Association, and especially his conclusion that “Whatever new comforts and luxuries may be provided in future by the advancement of physical science, it is on the development of the biological sciences that the peace and prosperity of the world will largely depend”. With this view we are in hearty agreement, though we should prefer to substitute for the word “largely” the word “ultimately”... But, if biology is to play the part required of it, it must be recognized as a profession, not as a hobby. It must be placed on the same professional level as chemistry, physics, engineering and medicine, with a corresponding prestige. For this reason we welcome the announcement of the proposal to establish an Institute of Biology.

From *Nature* 4 December 1948.

Solar system

Circular problems

Douglas P. Hamilton

The most striking regularity observed in our Solar System is that the planets all follow nearly circular, nearly co-planar orbits about the Sun. This feature lends strong support to the theory that the planets accreted out of the solar nebula — a flattened disk of gas and dust that surrounded the young Sun (Fig. 1). But do flattened disks of dust and gas inevitably lead to a few massive and widely spaced planets on circular uninclined orbits? Results from numerical simulations discussed at a meeting* in October suggest that, in as far as the inner Solar System is concerned, this may not be so.

The formation of the terrestrial planets — Mercury, Venus, Earth and Mars — can be loosely divided into three stages. Initially, dust grains move slowly through the solar nebula, bumping into one another in low velocity, 'sticking' collisions that gradually build up kilometre-sized 'planetesimals'¹. Next, gravity from the planetesimals begins to increase the accretion rate; this second stage is characterized by a rapid period of 'runaway growth' in which the largest planetesimals outcompete and ultimately absorb their smaller neighbours².

During these first two stages of planetary formation, gas drag from the solar nebula removes energy from the orbits of solid bodies, and damps away their radial and vertical motions as well. Thus solid bodies are thought to follow circular uninclined orbits that gradually evolve inward towards the Sun. The inward migration brings planetesimals together, and their nearly circular orbits lead to low relative velocities and sticking collisions, rather than to high-energy, destructive impacts.

This simple picture is altered by the gravitational perturbations of Jupiter and Saturn, which are traditionally thought to have formed after 10^6 – 10^7 years (ref. 3), roughly around the time that there were several dozen Mercury-sized 'protoplanets' in the terrestrial region. According to the standard

theory, the giant planets formed in a two-step process: first, solid cores of rock and ice accreted, and then the cores captured large quantities of gas from the solar nebula. Alternative theories, however, suggest that the giant planets formed directly from gas instabilities (without preceding core formation) in as little as 100–1,000 years (ref. 4).

New numerical simulations of this second stage of planetary formation have been carried out, working on the assumption that the shorter timescale is correct (S. Kortenkamp and G. Wetherill, Carnegie Inst., Washington, DC). These authors find that, when both gas drag and jovian perturbations are important, different-sized planetesimals follow slightly eccentric orbits which are out of phase with one another⁵. Collisions between these objects are often energetic enough to prevent accretion, thereby hindering the formation of protoplanets.

The magnitude of the effect depends sensitively on the orbits of the giant planets; it is diminished by the fact that Jupiter probably formed further from the Sun than it is now, subsequently migrating inward over 10^7 – 10^8 years (ref. 6). Nevertheless, if Jupiter formed quickly, this mechanism slows the runaway-growth timescales for the terrestrial planets, and significantly hinders the formation of large planetesimals in the asteroid belt. It also suggests that if large Jupiter-mass planets in other planetary systems formed quickly rather than slowly, then the zone in which Earth-like planets could form might be much reduced.

At the beginning of the third and final stage of terrestrial planet formation, the largest objects have attained sizes of small planets (about 3,000 km in diameter). Several dozen to several hundred of these protoplanets perturb one another gravitationally and merge in giant collisions until ultimately only the four terrestrial planets and the Moon remain. At the meeting, numerical simulations of this stage of planetary forma-

tion were reported by two teams — J. Chambers (Armagh Observ.)⁷; C. Agnor (Univ. Colorado), and R. Canup and H. Levison (Southwest Res. Inst., Boulder). These and other groups⁸ follow some 50 protoplanets for about 10^8 years. With plausible starting conditions, their models produce inner planetary systems with roughly the correct number, sizes and orbital spacings of planets. All groups, however, find that Earth-sized objects have orbital eccentricities and inclinations that are 5–10 times larger than those of Earth and Venus today. The eccentricities and inclinations are pumped up early in the simulations, primarily by secular interactions between the orbits of the protoplanets⁷.

So how did Venus and Earth end up on nearly circular orbits? There are only a few possibilities. First, the simulations may exclude some critical physics which damps orbital eccentricities and inclinations (for example dynamical friction with smaller objects, or larger-than-expected drag from nebular gas); this would cause planets on circular orbits to be formed more readily than the simulations suggest. Second, perhaps the inner planets in our Solar System did form with initially large eccentricities and inclinations, and these have been subsequently damped by some dissipative force (solar tides, for instance, although these are thought to be too weak) over the 4.5 billion years of Solar System history. Finally, perhaps elliptical orbits, rather than circular ones, are indeed the natural product of planetary formation. Using the anthropic principle, Chambers pointed out that if nearly circular orbits are necessary for climatic stability and the development of intelligent life, then Earth's low eccentricity can be understood. This argument is not particularly compelling, however, because it explains neither the low eccentricity of Venus, nor the low inclinations of Venus and Earth.

None of these three possibilities is completely satisfying, and so the circular problem remains. Its solution, which will have important ramifications for understanding our Solar System and extrasolar planetary systems, needs to be pursued. □

Douglas P. Hamilton is in the Department of Astronomy, Univ'ersity of Maryland, College Park, Maryland 20742, USA.

e-mail: hamilton@astro.umd.edu

1. Weidenschilling, S. J. & Cuzzi, J. N. *Protoplanets and Planets III* (eds Levy, E. H. & Lunine, J. I.) 1031–1060 (Univ. Arizona Press, Tucson, 1993).
2. Wetherill, G. W. & Stewart, G. R. *Icarus* **106**, 190–209 (1993).
3. Pollack, J. B., Hubickyj, O. & Greenzweig, Y. *Icarus* **124**, 62–85 (1996).
4. Boss, A. P. *Science* **276**, 1836–1839 (1997).
5. Kortenkamp, S. J. & Wetherill, G. W. *Icarus* (submitted).
6. Fernandez, J. A. & Ip, W. H. *Planet. Space Sci.* **44**, 431–439 (1996).
7. Chambers, J. E. & Wetherill, G. W. *Icarus* (submitted).
8. Ito, T. & Tanikawa, K. *Icarus* (submitted).

*Thirtieth Annual Meeting of the Division for Planetary Science (AAS), Madison, Wisconsin, 11–16 October 1998.



Figure 1 The solar nebula, by artist Don Davis. The gaps visible in the gaseous disk are swept clear by the largest protoplanets. One such object, perhaps destined to become Jupiter, is in the foreground.

Excitatory synapses

Is bigger better?

Robert C. Malenka and Roger A. Nicoll

The fundamental element of communication between neurons is the synapse, a subcellular specialization that requires the precise alignment of pre- and postsynaptic proteins. Over the past decade we have identified many of the molecular components involved in the exocytosis of neurotransmitter, but it has been more difficult to define the molecular architecture of postsynaptic sites, which differ in their protein composition depending on which neurotransmitter is used. A breakthrough in understanding excitatory synapses was the identification of a family of proteins that contain a PDZ domain. The properties of these proteins indicated that they may be important in the clustering and, perhaps, targeting of postsynaptic proteins^{1,2}. Now, on page 433 of this issue, Migaud *et al.*³ describe a mouse with a targeted mutation in a prominent PDZ-domain protein, postsynaptic density-95 (PSD-95; also known as SAP90⁵). Their findings confirm that PSD-95 is a key component of excitatory synapses,

but also indicate a function other than that originally envisaged.

Excitatory synapses in the mammalian brain contain concentrated clusters of the NMDA (*N*-methyl-D-aspartate) receptor, a subtype of glutamate receptor that is particularly important for triggering synaptic plasticity. Several years ago⁶, the cytoplasmic, carboxy-terminal tails of NMDA-receptor subunits were used as bait in a yeast two-hybrid screen and found to interact with PSD-95. This made perfect sense because PSD-95 was originally isolated⁴ as a prominent component of the postsynaptic density, — the electron-dense thickening found at all excitatory synapses. We now know that PSD-95 is one of a family of mammalian proteins that interact with NMDA-receptor subunits, as well as with certain potassium channels. This family also includes PSD-93/chapsyn-110, SAP97/Hdlg and SAP-102, all of which contain three tandem PDZ domains at the amino terminus, an SH3 (src-homology) domain and a carboxy-terminal domain

homologous to yeast guanylate kinase^{1,2}. PDZ domains are thought to mediate protein–protein interactions and, in the case of the PSD-95 family, they bind to short amino-acid motifs at the carboxy termini of interacting proteins^{1,2}.

Several lines of evidence support the idea that PSD-95 is involved in the synaptic clustering and targeting of NMDA receptors. First, both are highly concentrated at synapses⁶. Second, during development, PSD-95 clusters form opposite presynaptic terminals several days before synaptic NMDA clusters appear⁷. Third, when PSD-95 or PSD-93/chapsyn-110 are coexpressed with NMDA receptors in heterologous cells (where these proteins are not normally expressed), clusters appear that do not form when the proteins are expressed alone⁸. Finally, mutations in the *Drosophila* homologue of PSD-95, Dlg, abolish the synaptic clustering of Shaker potassium channels⁹.

Although highly suggestive, all of this evidence is indirect, so neurobiologists have eagerly anticipated the first report of mice that lack PSD-95. This feat has now been accomplished by Migaud *et al.*³. They placed a stop codon in the third PDZ domain of PSD-95, leaving only the first two domains intact. The first surprising finding is that,

Ocean chemistry

Iron and brimstone

Spectacular layered sedimentary rocks, such as those pictured here from the Hamersley Basin of Western Australia, are a widespread yet episodic feature of the Precambrian geological landscape. These 'banded iron formations' — the colouring is largely due to variable iron content — were laid down from an iron-rich ocean. Periods during which they accumulated are generally thought to have been ended by oxygenation of sub-surface ocean waters, forcing the oxidation and almost complete precipitation and deposition of dissolved iron from hitherto anoxic waters. However, as D. E. Canfield reports elsewhere in this issue (*Nature* 396, 450–453; 1998), such a deep-reaching breath of fresh air may not, after all, have been needed to relieve the ocean of its ferrous burden.

Canfield bases his case on a simple model of cycling of oceanic oxygen and nutrients, along with measured sulphur-isotope signatures of sedimentary sulphide deposits and evidence of the emergence of various life forms. He argues that the ocean could have lost its iron from bottom waters that were akin to an oxygen-free sulphide 'soup'. Focusing on the Proterozoic Eon (from about 2.5 billion years ago to the end of the Precambrian around 540 million years ago), he proposes that known



episodes of oxidation of the Earth's surface were insufficient to make the deep ocean aerobic until at least one billion years ago. If so, the end of the last great deposition of banded iron formations, some 1.8 billion years ago, stemmed not from the oxygenation of the deep — as generally held — but from some other process.

The sulphur-isotope data implicate a shift in the marine sulphur cycle as an alternative explanation. Canfield hypothesizes that a known oxidation of the Earth's surface about two billion years ago increased seawater sulphate

concentrations by oxidizing continental mineral sulphides. Such an increase could have relieved existing limitations on the oceanic rate of sulphate reduction to sulphide, so allowing sufficient build-up of sulphide to precipitate out the ocean's dissolved iron as iron sulphides. On this analysis, the great oxygenation of the ocean interior did not occur for about another billion years, following another Earth-oxidation event which swept the deep waters clean of sulphide and, perhaps literally, breathed new life into the ocean.

Philip Newton

D. E. CANFIELD

although PSD-95 is no longer found at synapses in the knockout mice, NMDA receptors are concentrated appropriately and generate normal synaptic responses. Thus, PSD-95 is not required for the synaptic clustering of NMDA receptors. Of course, other members of the PSD-95 family, such as PSD-93/chapsyn-110, may have served this role.

The second surprising finding comes from the study of synaptic plasticity in the CA1 region of the hippocampus. Normally, depending on the specific pattern of synaptic activation¹⁰, CA1 synapses show both long-term potentiation (LTP) and long-term depression (LTD). In the absence of PSD-95, not only did Migaud and colleagues find that LTP was of much greater magnitude than normal, but the synaptic activity that normally elicits LTD now generated robust LTP. In other words, the absence of PSD-95 disrupted the rules by which synaptic activity is translated into long-lasting changes in synaptic strength. Presumably, this alteration explains why the mutant mice were considerably impaired in a hippocampal-dependent learning task.

Why were these synaptic-plasticity mechanisms disrupted? As well as NMDA receptors^{6,8} PSD-95 can interact with other proteins, some of which are involved in signalling¹¹. So, like other PDZ-domain proteins^{12,11,12}, PSD-95 has been proposed to act as a scaffold that couples a signalling complex to activation of NMDA receptors. Presumably, without PSD-95 some component of the signalling complex that constrains LTP and triggers LTD has been lost. What might this be? One model for the bidirectional control of synaptic strength¹⁰ involves a tug-of-war between protein kinase (LTP) and protein phosphatase activity (LTD), so perhaps a critical protein phosphatase in the post-

synaptic density is missing. To date, however, there is no evidence for an interaction between phosphatases and PSD-95.

Another intriguing finding is that the PSD-95 mutant mice showed an increase in paired-pulse facilitation — a short-lived, activity-dependent increase in neurotransmitter release. This suggests that the lack of postsynaptic PSD-95 has influenced presynaptic function, perhaps by disrupting the interaction with other proteins (such as a recently identified presynaptic tripartite complex¹³).

Migaud and colleagues³ elegantly show that we now have the tools to understand the molecular architecture of excitatory synapses in the mammalian brain. Their study also emphasizes the surprises that are in store as we try to work out how a specific protein embedded in this complex architecture leads to synaptic function and, eventually, to behaviour. □

Robert C. Malenka is in the Departments of Psychiatry and Physiology, and Roger A. Nicoll is in the Departments of Cellular and Molecular Physiology and Physiology, University of California, San Francisco, California 94143, USA.

e-mails: malenka@itsa.ucsf.edu

nicoll@phy.ucsf.edu

- Sheng, M. *Neuron* **17**, 575–578 (1996).
- Kornau, H.-C., Seeburg, P. H. & Kennedy, M. B. *Curr. Opin. Neurobiol.* **7**, 368–373 (1997).
- Migaud, M. *et al. Nature* **396**, 433–439 (1998).
- Choi, K.-H., Hunt, C. A. & Kennedy, M. B. *Neuron* **9**, 929–942 (1992).
- Kistner, V. *et al. J. Biol. Chem.* **268**, 4580–4583 (1993).
- Kornau, H.-C., Schenker, L. T., Kennedy, M. B. & Seeburg, P. H. *Science* **269**, 1737–1740 (1995).
- Rao, A., Kim, E., Sheng, M. & Craig, A. M. *J. Neurosci.* **18**, 1217–1229 (1998).
- Kim, E., Choi, K.-O., Rothschild, A. & Sheng, M. *Neuron* **17**, 103–113 (1996).
- Tejedor, F. J. *et al. J. Neurosci.* **17**, 152–159 (1997).
- Malenka, R. C. *Cell* **78**, 535–538 (1994).
- Craven, S. E. & Bredt, D. S. *Cell* **93**, 495–498 (1998).
- Tsunoda, S. *et al. Nature* **388**, 243–249 (1997).
- Butz, S., Okamoto, M. & Südhof, T. C. *Cell* **94**, 773–782 (1998).

Optics

A photonic crystal fibre

Pauline Rigby

Ordinary optical fibres guide light by total internal reflection, which relies on the refractive index of the central core being greater than that of the surrounding cladding. Writing in *Science*, Jonathan Knight and colleagues¹ describe a fundamentally different type of optical fibre, one that has a core with a lower refractive index than the cladding and so rules out the possibility of total internal reflection. Instead, light is guided by a mechanism which allows it to be piped through air.

The core of the new fibre is essentially a defect in an otherwise periodic array of air holes placed at the vertices of a honeycomb lattice (Fig. 1). This fibre is an example of a photonic crystal — a structure that repeats in

one or more directions in space^{2,3}. As a result of multiple reflections, certain wavelengths cannot propagate in these structures. A structure is said to possess a 'photonic bandgap' if it reflects a wavelength incident from any angle in space.

In the early days of this field, it seemed that a large contrast in refractive index, such as that between air and a semiconductor (which has a refractive index, n , greater than 3), would be needed to achieve a photonic bandgap. But Birks *et al.*⁴ showed in theory that it is possible to create a two-dimensional photonic bandgap using the modest contrast in refractive index between air and silica ($n = 1.5$), providing that the light has a component travelling parallel to the direction of the rods.

Adding an extra air hole, or removing a hole, creates a core through which light is guided. Knight *et al.*⁵ previously made a photonic crystal fibre with a missing air hole at its centre to form a solid core. This fibre has some remarkable properties — however large the core, it supports only one optical mode (an intensity distribution of light that remains constant as it travels along the fibre)⁶. This type of fibre would also transmit very high optical powers in comparison to conventional fibres, not because the mode is in air but because the area containing the

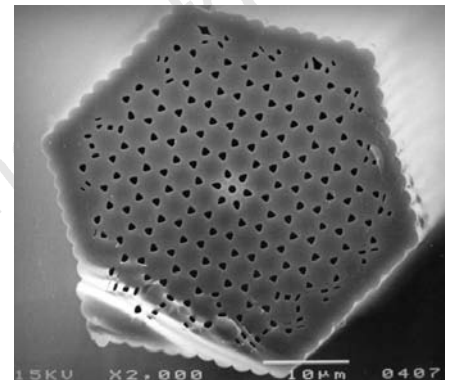


Figure 1 Cross-section of the photonic crystal fibre of Knight *et al.*¹. The fibre is made by stacking a few hundred solid silica rods and hollow capillary tubes (which create the air holes) in a close-packed hexagonal pattern to make a preform, which is then drawn on a fibre-pulling tower at a temperature of about 2,000 °C. The lateral dimensions shrink by a factor of around 10,000, while becoming extended in the third direction. The holes run through the length of the fibre.

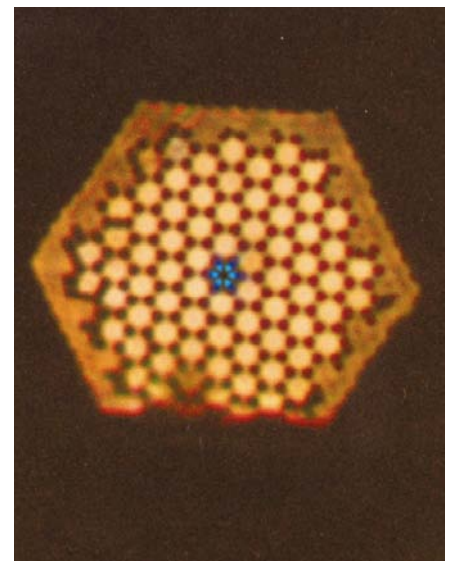


Figure 2 Guided light. In the photonic crystal fibre, the shape of the guided mode (blue) is influenced by six small holes about the main core (see Fig. 1). These survive near the centre of the fibre, but have collapsed elsewhere because of the forces introduced by the central hole.

light could be made very large⁷. The fibre remains single-mode as long as the ratio of air-hole-size to spacing is kept constant⁶. Knight *et al.* are convinced, however, that the guiding mechanism in this fibre with a solid-silica core is still total internal reflection. The outer part of the fibre behaves as an 'effective index' medium, in which the average refractive index is reduced by the presence of the air holes.

The key feature of the new photonic bandgap fibre¹ is that it guides light in air (Fig. 2). This could have some potentially important applications. For example, wavelengths that would be absorbed by silica could be transmitted for much longer distances through air; and because air is not susceptible to the nonlinear effects that occur in silica at moderate optical powers, much higher powers could be delivered.

Another property of the fibre is strong birefringence, meaning that two orthogonal modes travel down the fibre at different velocities. We expect this intuitively because the modes do not 'see' the same pattern of holes: the fibre does not appear the same when turned through 90°. In a strongly birefringent fibre it is difficult to couple one mode into the other. As a result, light in the fibre will remain linearly polarized even if the fibre is bent or twisted, a property that could be exploited in polarization-preserving fibres.

Several difficulties must be surmounted to achieve absolute command over the design and production of photonic crystal fibres. On the theoretical side, Knight *et al.* confess that they do not fully understand the origin of the birefringence; nor have they been able to predict exactly which wavelengths will be guided because the model only caters for perfectly round holes. Practical problems include determining the bending and transmission losses of the fibre, as well as making it in useful lengths and at any scale. Nonetheless, these issues are well worth tackling — photonic crystal technology offers opportunities for designing fibres with features that are unattainable using conventional methods. □

Pauline Rigby is in the Department of Engineering Science, University of Oxford, Parks Road, Oxford OX1 3PJ, UK.

e-mail: pauline.rigby@eng.ox.ac.uk

1. Knight, J. C., Broeng, J., Birks, T. A. & Russell, P. St J. *Science* **282**, 1476–1478 (1998).
2. Joannopoulos, J. D., Meade, R. D. & Winn, J. N. *Photonic Crystals: Molding the Flow of Light* (Princeton Univ. Press, 1995).
3. Cassagne, D., Jouanin, C. & Bertho, D. *Phys. Rev. B* **52**, R2216–R2220 (1995).
4. Birks, T. A., Roberts, P. J., Russell, P. St J., Atkin, D. M. & Shepherd, T. J. *Electron. Lett.* **31**, 1941–1943 (1995).
5. Knight, J. C., Birks, T. A., Russell, P. St J. & Atkin, D. M. *Optics Lett.* **21**, 1547–1549 (1996).
6. Birks, T. A., Knight, J. C. & Russell, P. St J. *Optics Lett.* **22**, 961–963 (1997).
7. Knight, J. C., Birks, T. A., Cregan, R. F., Russell, P. St J. & Sandro, J. P. *Electron. Lett.* **34**, 1347–1348 (1998).

Signal transduction

New exchange, new target

Julian Downward

The Ras superfamily of small guanine-nucleotide-binding (G) proteins controls many aspects of cellular behaviour, from proliferation and survival to morphology and membrane transport. Despite their varied effects, these proteins are regulated by a common mechanism — the opposing effects of stimulatory guanine-nucleotide-exchange factors (GEFs), which promote the uptake of fresh GTP to form an active, GTP-bound conformation, and inhibitory guanosine triphosphatase activating proteins, which stimulate hydrolysis of bound GTP to form the inactive, GDP-bound conformation.

On page 474 of this issue, de Rooij *et al.*¹ report a new twist to the mechanism for control of the Ras-related protein Rap1. Their study shows that regulation of proteins in the Ras superfamily is considerably more varied than previously thought, and establishes, for the first time, the molecular details of how a Ras-related protein is controlled. As well as all this, the authors identify a new class of tar-

get protein for cyclic AMP, Earl Sutherland's original second messenger molecule, raising the possibility that several of cAMP's effects on cells may not be mediated by its previously characterized targets.

When cells are treated with a growth factor, Ras is activated by its GEF, Sos. A complex forms containing the growth-factor receptor (an autophosphorylated tyrosine kinase), adaptor proteins with SH2 and SH3 (src-homology) domains, and Sos². Assembly of this complex is not thought to stimulate Sos directly; instead, translocation of Sos from a cytosolic location to the plasma membrane — where both the receptor and Ras are localized — is believed to give Sos improved access to Ras (Fig. 1a). However, there has been no totally convincing proof or *in vitro* reconstitution of this model.

It is also clear that other Ras GEFs exist, and that they are regulated quite differently from Sos. One, for example, the Ras guanine-nucleotide-releasing factor (Ras-GRF), is stimulated by the direct binding of

Ca²⁺-calmodulin³ and also by the action of heterotrimeric G proteins⁴. Earlier this year another GEF, Ras guanine-nucleotide-releasing protein (RasGRP), was identified in the brain⁵. RasGRP contains Ca²⁺-binding motifs and a diacylglycerol-binding domain, and it is activated on treating cells with diacylglycerol — possibly because it is then relocalized to the plasma membrane.

Rap1 is a close and, until fairly recently, neglected relative of Ras. It was identified in a screen⁶ for molecules that inhibit the transforming function of Ras. Although Rap1 can interact with the same effectors as Ras, it is localized to intracellular membranes — this may explain why it can antagonize Ras when overexpressed. There is conflicting evidence as to whether its ability to antagonize Ras is physiologically relevant (reviewed in ref. 7), although the balance seems to be shifting away from a physiological role⁸. There are also data to support an alternative model for Rap1 function in which it has *similar* effects to Ras, particularly the ability to activate B-Raf and, hence, the mitogen-activated protein (MAP) kinase cascade⁹. But, in most systems Rap1 has very different effects to Ras so, although they may share some signalling pathways, Rap1 probably activates these systems in distinct ways and may also interact with unique effectors⁷.

Rap1 is activated in response to a range of stimuli through a number of second messenger molecules, including cAMP, Ca²⁺ and diacylglycerol^{8,10,11}. The only previously identified GEF for Rap1 is C3G, which seems to function analogously to Sos. But it is unlikely that C3G mediates activation of Rap1 by cAMP, and de Rooij *et al.*¹ now show that another GEF is responsible.

Protein kinase A (PKA) is activated by cAMP, so we could imagine a mechanism whereby cAMP activates PKA which, in turn, leads to activation of Rap1. But, by using inhibitors and mutants, de Rooij *et al.* found that PKA is not involved in the cAMP-induced activation of Rap1. Indeed, although PKA phosphorylates Rap1 near its carboxy terminus¹⁰, this phosphorylation is not required for cAMP-dependent activation. This led the authors to search sequence databases for possible Rap1 GEFs that might be directly regulated by cAMP, and they pulled out a protein which they named Epac (exchange protein directly activated by cAMP). As well as containing homologies to other GEFs for Ras-like proteins, Epac possesses sequences related to the regulatory subunit of PKA. De Rooij and colleagues found that Epac can bind directly to cAMP, and that its activity towards Rap1 is stimulated if the levels of cAMP in the cell are increased (Fig. 1b). Epac is, therefore, not only another GEF for Rap1, but an example of a new type of cAMP sensor protein. It is only the third such type identified to date,

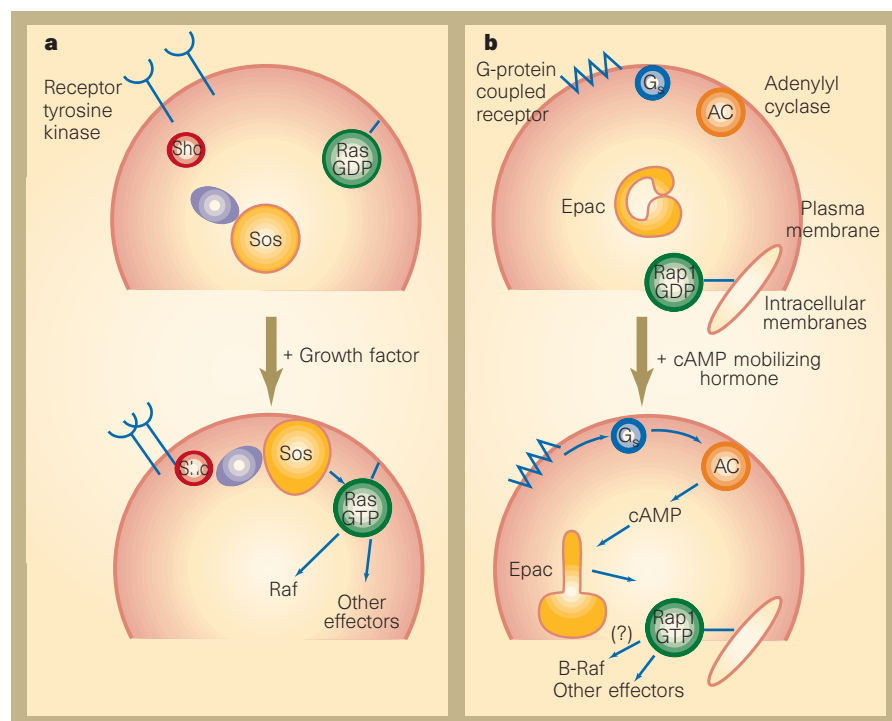


Figure 1 Two types of regulatory system for Ras proteins. **a**, The classic Sos pathway. Activation of a growth-factor receptor tyrosine kinase leads to formation of complexes containing adaptor proteins (Shc and Grb2) and Sos (a guanine-nucleotide-exchange factor; GEF), at the plasma membrane where Ras is also located. Increased local concentration of Sos results in nucleotide exchange on Ras to form the active GTP-bound conformation, which stimulates several effector enzymes including Raf. **b**, Increases in the concentration of cyclic AMP, caused *in vivo* by hormone activation of G_i-coupled receptors, results in the binding of cAMP to Epac, a Rap1-specific GEF identified by de Rooij *et al.*¹. This causes a conformational change leading to increased exchange activity towards Rap1, which is localized on intracellular membranes. GTP-bound Rap1 then activates effectors that may include B-Raf.

after cAMP-dependent protein kinases and cyclic-nucleotide-gated ion channels.

Epac seems to fit into the category of second-messenger-stimulated GEFs (such as Ras-GRF and RasGRP), rather than the adaptor-regulated GEFs like Sos and C3G. However, this new work goes well beyond any previously done on Ras-family GEFs because de Rooij *et al.* used purified components to show, *in vitro*, that the activity of Epac towards Rap1 is allosterically stimulated on binding cAMP. They also deleted the cAMP-binding domain of Epac and found that this led to activation of Epac *in vitro*. This indicates that the cAMP-binding domain normally inhibits the exchange activity of Epac until cAMP binds, most likely causing a conformational change that relieves the inhibition.

This is the first time that the regulation of a Ras-family GEF has been reconstituted with defined components, and de Rooij *et al.* provide mechanistic detail at the molecular level that is still lacking for much more intensively studied GEFs such as Sos. After years spent languishing in the shadow of its flamboyant cousin, Rap1 has finally stepped into the limelight itself. Although there may still be disagreement as to the downstream effects of Rap1 on cell function, it is clearly a broadly

used sensor for second messengers such as cAMP, Ca²⁺ and diacylglycerol. Moreover, the localization of Rap1 on intracellular membranes, which are readily accessible to soluble second messengers, may provide further clues to its function. □

Julian Downward is in the Signal Transduction Laboratory, Imperial Cancer Research Fund, 44 Lincoln's Inn Fields, London WC2A 3PX, UK. e-mail: downward@icrf.icnet.uk

1. de Rooij, J. *et al.* *Nature* **396**, 474–477 (1998).
2. McCormick, F. *Nature* **363**, 15–16 (1993).
3. Farnsworth, C. L. *et al.* *Nature* **376**, 524–527 (1995).
4. Mattingly, R. R. & Macara, I. G. *Nature* **382**, 268–272 (1996).
5. Ebinu, J. O. *et al.* *Science* **280**, 1082–1086 (1998).
6. Kitayama, H., Sugimoto, Y., Matsuzaki, T., Ikawa, Y. & Noda, M. *Cell* **56**, 77–84 (1989).
7. Bos, J. L. *EMBO J.* **17**, 6776–6782 (1998).
8. Zwartkruis, F. J., Wolthuis, R. M., Nabben, N. M., Franke, B. & Bos, J. L. *EMBO J.* **17**, 5905–5912 (1998).
9. York, R. D. *et al.* *Nature* **392**, 622–626 (1998).
10. Altschuler, D. L., Peterson, S. N., Ostrowski, M. C. & Lapetina, E. G. *J. Biol. Chem.* **270**, 10373–10376 (1995).
11. Franke, B., Akkerman, J. W. & Bos, J. L. *EMBO J.* **16**, 252–259 (1997).

erratum

In the caption to Figure 1 of “Strategies for cutting carbon” (D. G. Victor, *Nature* **395**, 837–838; 1998) the energy unit should be terawatt years, not terawatt hours. □

Daedalus

Watch this space

Empty space is a hive of activity. Particle–antiparticle pairs are being created out of nothing, all the time and everywhere, and vanishing again before their energy can violate the uncertainty principle. A high-energy particle collision nearby may provide the energy to stabilize such a virtual pair, making it real.

What velocity, asks Daedalus, is a virtual pair born with? The obvious answer is that it is created stationary; but relativity makes this impossible. Empty space has no reference frame to be stationary relative to. The best it can do, says Daedalus, is to arrive with some random velocity relative to local matter. And a truly random velocity, somewhere between plus infinity and minus infinity, is almost certain to be very large indeed.

Indeed, such pairs might constitute the long-sought tachyons, which travel faster than light and backwards in time. But Daedalus reckons that to external observers, new pairs will merely seem to be moving extremely close to the speed of light. Relativistic time-dilation will therefore stretch their internal timescale enormously. Even if a pair cannot last more than 10^{−20} or 10^{−30} seconds before vanishing again, to external observers the time will seem much longer. At almost the speed of light, the pair could travel some distance before disappearing. It could even collide with something, and (just as in a violent particle collision) be transformed into something detectable.

Daedalus would like to test this bold theory. Several large-scale experiments, such as neutrino telescopes and proton-decay detectors, consist of a large volume full of something, surrounded by a dense array of detectors all looking in. Daedalus proposes to pump out such an experiment, leaving the detectors watching a vacuum. Any ray, particle or pair that enters from outside will collide with and trigger two detectors, at its entry and exit points. But one created within the vacuum will trigger only one.

A successful outcome would provide all sorts of new high-energy events. If it detects newly created hydrogen atoms, it would support the ‘continuous creation’ theory, in which a steady-state Universe is maintained by the continuous appearance of new hydrogen throughout space. And if it sees only pairs with a well-defined velocity with respect to local matter, it will disprove relativity and resurrect that universal inertial reference frame, the ether.

David Jones

Obituary

Jürgen Aschoff (1913–98)

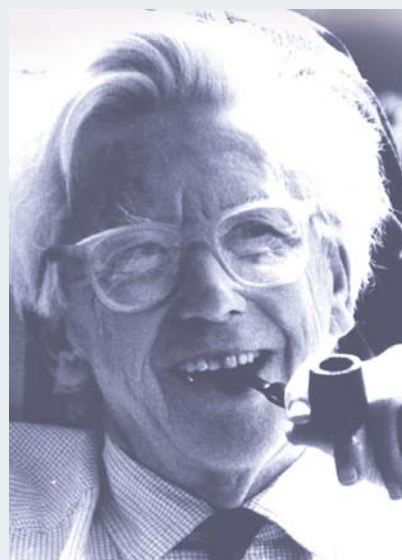
Pioneer in biological rhythms

For chronobiology, an era came to an end with the death of Jürgen Aschoff on 12 October at the age of 85. Along with his younger friend Colin Pittendrigh, whom he survived by two years (see *Nature* 381, 24; 1996), Aschoff was regarded as the founder of the study of biological rhythms. Together they laid out the conceptual framework for our current understanding of, and vigorous enquiry into, the temporal organization of life as it has evolved on a rotating planet.

The youngest son of a famous father — the Freiburg pathologist Ludwig Aschoff — Jürgen grew up in the liberal but morally strict milieu of Prussian academia. He studied medicine at the University of Bonn, then moved to the University of Göttingen where his scientific career began. There, his teacher Hermann Rein worked on thermoregulation physiology, inspiring Aschoff to follow in his footsteps (he was able to keep Aschoff out of the war, as indispensable for the training of physicians). After the war, Aschoff lectured at the University of Würzburg, and a gifted teacher he turned out to be.

In 1952 Rein was appointed director at the new Max Planck Institute for Medical Research in Heidelberg. He took his brightest student with him as a collaborator. Experimenting on himself, Aschoff discovered spontaneous 24-hour variations in human heat loss. This aroused his interest in the mechanism behind such patterns. It was known that activity rhythms in animals persisted in constant conditions and were endogenously generated. Raising birds from the egg, and breeding mice for several generations in constant conditions, Aschoff established that this rhythmicity was innate: no previous exposure to the 24-hour day was required to evoke it.

In the late 1950s Konrad Lorenz and Erich von Holst lured Aschoff to their new Max Planck Institute of Behavioural Physiology (MPIV) in Bavaria. There he built his own department dedicated to the study of biological timing. As a physician among behavioural biologists, he grasped the opportunity of studying humans and animals in parallel. In birds and mammals, he investigated the sensitivity of endogenous circadian systems to environmental stimuli. The intensity of continuous illumination modulated their frequency in predictable though opposite ways in nocturnal and diurnal animals, a



phenomenon soon generalized as Aschoff's rule.

These and other findings led to a new conceptual view of the synchronization of circadian rhythms. It postulated an innate biological oscillator which under natural conditions is synchronized with the Earth's rotation by the response to a 'zeitgeber' — a word Aschoff coined and contributed to the English language. Applying physical oscillator theory, Aschoff was able to make and experimentally test predictions about the behaviour of circadian systems in response to different zeitgeber properties. His experimental and theoretical work in the 1950s and 1960s laid the basis for viewing circadian rhythms as the product of endogenous oscillators which derive their true functional significance from the maintenance of a constant phase relationship with the light–dark cycle — the most precise zeitgeber the planet has to offer. The demonstration of localized light-sensitive circadian pacemakers by others in the 1970s confirmed the validity of the approach.

Meanwhile, Aschoff went on to apply the same experimental approach to humans. He built an underground 'bunker', designed to study human subjects in prolonged isolation from time cues. Characteristically, he chose to be the first subject himself. In a series of investigations over 20 years he and his collaborator Rütger Wever firmly established that human physiology, behaviour and cognition are just as much controlled by endogenous 'circadian' oscillators as those of our fellow creatures. These findings had far-reaching effects on the medical sciences — arguably more fundamentally so than the work of Aschoff's influential father. It

led to our current understanding of many socio-medical problems that stem from shift work, affective disorders, sleep disorders, ageing and jet lag. It also provided the basis for optimizing drug-application schedules in pharmacological therapy.

The 'Abteilung Aschoff' at the MPIV in the picturesque village of Andechs exerted a magnetic attraction on researchers from a wide variety of backgrounds. Biologists, psychologists, mathematicians and physicians arrived to discuss their research. They learned to talk about biological rhythms in a common language. Aschoff had an unusual breadth of mind, a clear focus on basic principles and an encyclopaedic knowledge of diverse literatures. In some of his reviews, he laid the groundwork for emerging fields. For instance, he contributed the core-shell concept to thermoregulation; as early as 1955, he forecasted the existence of endogenous circannual timers in annual reproduction; and he analysed human seasonal conception patterns and introduced the idea of a 'biological equator'.

Like Pittendrigh, Aschoff had a great talent for lecturing. He combined scientific creativity with *joie de vivre* and a warm personal interest in his fellow scientists. The atmosphere so created made a pilgrimage to the Mecca of biological rhythms research a lasting source of inspiration to scores of scientists.

After retiring in 1983, Aschoff returned to live in the stately mansion in which he was born, on the Ludwig Aschoff Platz in Freiburg. He maintained a continuous stream of original publications — including many on the human sense of time — several of which are still in press. Aschoff's exuberant vitality was eventually broken when his lifelong partner Hilde Jung died, ten months ahead of him. She too was a warm and unconventional personality, who had a major share in creating the environment in which he flourished. They leave six children, 19 grandchildren and an uncountable number of scientific offspring. With his daring departures from established disciplines, Aschoff set an unforgettable example to those who had the privilege to know him and learn from him.

Serge Daan and Eberhard Gwinner

Serge Daan is in the Zoological Laboratory, University of Groningen, PO Box 14, 9750 AA Haren, The Netherlands.

e-mail: S.Daan@biol.rug.nl

Eberhard Gwinner is in the Forschungsstelle für Ornithologie der Max-Planck-Gesellschaft, von der-Tann-Strasse 7, 82346 Andechs, Germany. e-mail: gwinner@erl.ornithol.mpg.de

Julesz's joyfulness

British 'natural magicians' of the nineteenth century could turn two flat images into one three-dimensional form. Later, Bela Julesz believed his stereograms showed how the brain turns images from two eyes into one reality.

Martin Kemp

Why two eyes result in one mental image has perplexed investigators of vision from the very beginning. However effective the techniques forged by artists and photographers to evoke three dimensions on a flat surface, the absence of parallax cues remained a serious limitation. It is surprising that the idea of presenting a slightly different image of the same object to each eye did not occur to the Jesuit 'magicians' of the Baroque era. In the event it was British exponents of 'natural magic' in the 1830s who performed the trick.

It is unclear whether credit for this should be given to James Elliot of Edinburgh in 1834, as claimed by the disputatious David Brewster, or to Charles Wheatstone, whose 1838 publication backdated his invention by at least six years. But it was Wheatstone's two-mirror stereoscope that entered the public arena. His earliest images, linear drawings of

geometrical solids which lacked other spatial clues, had already shown how fundamental the process of stereopsis was in the presentation of "a solid figure... in such a manner that no effort of imagination can make it appear as a representation on a plane surface".

The making and viewing of stereo-images was progressively refined over the years, but the next fundamental move came with the invention of the random-dot stereogram in its definitive, computerized form by Bela Julesz in 1959. He showed how the "cyclops within us" exploits stereo processing to unscramble two apparently unintelligible random-dot arrays in which a central region is invisibly 'lifted' and displaced to the right in the left image and vice versa. Using an appropriate viewing device, or crossing our eyes in front of the images, we discern the displaced zone as if it were a shape suspended in front of the background.

Looking back on Julesz's 1971 book, *Foundations of Cyclopean Perception*, it is

striking how much it stands in line of succession from Brewster's *Letters on Natural Magic* (1832) and his monographs on the kaleidoscope and stereoscope, and how different in tone it is from David Marr's posthumously published *Vision* in 1982, in which a "cooperative algorithm" is posited as a computational alternative to Julesz's more mechanistic explanatory framework.

Julesz, like Brewster, stands within a tradition of culturally expansive science. Not only does Julesz joyfully make big claims for the technical innovation as the gateway to a theory of "global stereopsis" and as a "real paradigm" but he also aspires to place his discovery on a "broad foundation by using analogies from many disciplines, including music". Characteristically, he calls his technique "random counterpoint".

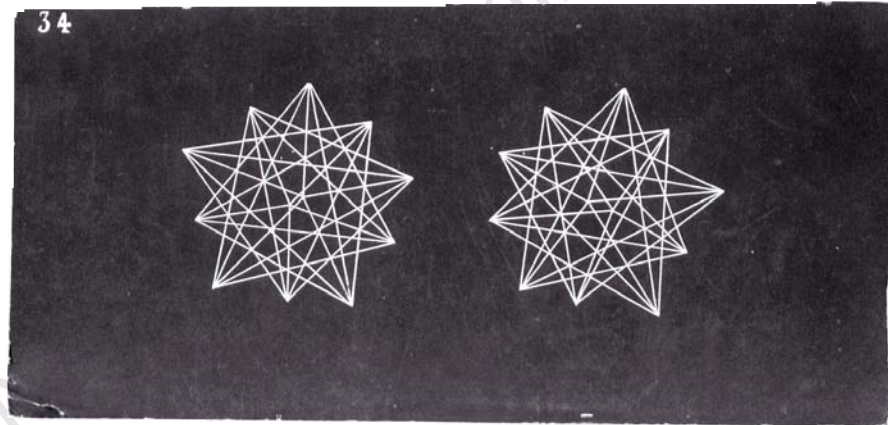
He claimed that random-dot stereograms allow experimenters to communicate directly with internal processing. This access to "the mind's retina" bypasses the peripheral processes in a way that "eclipses" monocular clues and isolates stereopsis "without rivalry". He proposed that, if the classic experiments on perception and illusion could be replicated in random-dot stereograms, the processes would be identifiable as centrally located within the brain rather than attributable to the retina.

His intended audience comprised, primarily, "students" and "workers in visual perception", together with "clinicians" who might want tests for stereopsis, and "the mathematician and designer" interested in "the visualization of complex surfaces". He also stressed that his illusions would, "last, but not least, delight the layman or student of the visual arts".

It is not my intention here to present a historical choreography of the successive refinements and interactions between the experimental 'grit' and the algorithmic modelling. Rather, I offer a reminder of how the experimental psychologists' modern bag of tricks, however esoteric the techniques and theories, can still serve to induce popular awe at the 'natural magic' residing with our perceptual system — with a sense of joy that is undiminished from the age that saw the invention of the stereoscope. □

Martin Kemp is in the Department of the History of Art, University of Oxford, 35 Beaumont Street, Oxford OX1 2PG, UK.

Quotations from Julesz are taken from B. Julesz, *Foundations of Cyclopean Perception* (Univ. Chicago Press, 1971).



Early stereo card of a stellated octahedron, for use with a lenticular stereoscope.



Get in touch with your inner cyclops: crossing the eyes should reveal a diamond hovering above the textured background. Julesz believed these random-dot stereograms gave researchers direct access to "the mind's retina" (from B. Julesz, *Foundations of Cyclopean Perception*).

Marine incursion into South America

The Amazon basin harbours the most diverse assemblage of freshwater fishes in the world, including a disproportionately large number of marine-derived groups, such as stingrays, flatfishes, pufferfishes, and anchovies¹. On the basis of our molecular phylogenetic analysis of South American freshwater stingrays (Potamotrygonidae), coupled with reconstructions of Amazonian palaeogeography, we propose that some marine-derived freshwater fish species originated as a by-product of massive movements of marine waters into the upper Amazon region during the Early Miocene epoch, 15–23 million years ago.

Miocene South America experienced profound changes of topography, environment and river drainage patterns^{2,3}. A combination of sea-level changes and tectonic loading of the foreland basin in the upper Amazon produced significant incursions of sea water, as indicated by the presence of marine and brackish mollusc, copepod and mangrove fossils^{4,5}. Nuttall⁴ says the upper Amazon during the Miocene was “up to 500 km wide, occupied by a continually shifting pattern of streams, swamps, and lakes of varying salinity and offering intermittent connections with the Caribbean”. These conditions would have been ideal for the isolation of marine fishes in progressively desalinized habitats. We suggest, therefore: first, that the divergence between potamotrygonid stingrays and their closest marine relative should have occurred during the Miocene; and second, that the distribution of the closest marine relative should include the Caribbean, a proposed source of the marine incursion.

To test the marine-incursion hypothesis, we investigated the evolutionary history of stingrays using DNA sequences from the gene encoding mitochondrial cytochrome *b* (Fig. 1a). We analysed sequences from ten South American freshwater stingrays, including representatives of all three potamotrygonid genera, and from all the supposed marine relatives of potamotrygonids. We calibrated a rate for cytochrome *b* evolution in potamotrygonids based on the amount of sequence divergence between species separated by a known geological event, the uplift of both the eastern cordillera of the Andes and the Merida Andes, which isolated the Maracaibo stingray (*Potamotrygon yepzei*) approximately 8 million years ago^{2,3}.

Based on our cytochrome *b* calibration and phylogenetic tree, the estimated divergence time between potamotrygonids and their closest marine relative indicates that neotropical freshwater stingrays originated in the Early Miocene. Moreover, one of the two species that are the marine relatives of

the freshwater rays is distributed along the northern coast of South America (Fig. 1b). Both the estimated origination time of the freshwater species and the biogeography of the closest marine relative are therefore consistent with predictions of the marine-incursion hypothesis.

Alternative explanations have been advanced to account for the presence of marine-derived taxa in South America. For instance, it has been suggested¹ that the extensive interface between marine and freshwater habitats in the lower Amazon served as a gateway for marine invaders of neotropical rivers. Such invasion hypotheses are difficult to refute because they are compatible with many different biogeographical patterns. Invasions are essentially opportunistic, and we might expect fresh water to have been colonized several times; however,

even though South American rivers have been open to the oceans for more than 100 million years, endemic freshwater stingrays have a single, unique origin. We believe that this pattern argues against invasions by marine rays, and suggests instead that a single geological event, such as a marine incursion, was crucial.

It has also been proposed that uplift of the Andes blocked a previously westward-draining proto-Amazon, and isolated a component of the Pacific marine fauna in progressively desalinized inland lakes or seas. This would mean a Late Cretaceous or Palaeocene origin for Potamotrygonidae, as this is when Andean uplift probably severed the last connection between the Pacific Ocean and the upper Amazon⁹. But we estimate that potamotrygonids originated more recently, and the marine relative of

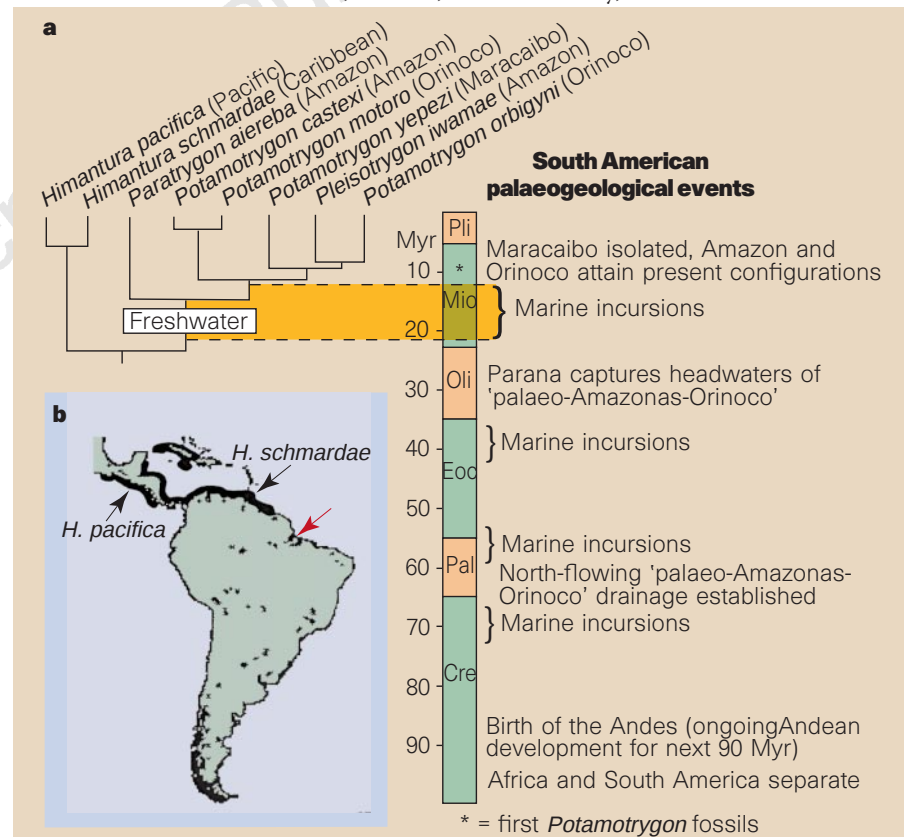


Figure 1 Phylogeny and biogeography of freshwater stingrays. **a**, Phylogenetic hypothesis for freshwater stingrays and their closest marine relative, and timescale showing palaeogeological events in South America^{2,3}. Relationships of Potamotrygonidae, ampho-American *Himantura* and seven other stingray genera were inferred from parsimony analysis of 765 base pairs of cytochrome *b* sequence, with transitions at third codon positions weighted one-fifth of all other changes¹⁰ (not all taxa shown; sequences available from Genbank). The potamotrygonid species are a monophyletic group, so neotropical freshwater rays had a single origin. *H. pacifica* and *H. schmardae* were supported as the closest sister taxa to potamotrygonids, consistent with morphological analysis⁹. Ages of nodes were estimated from transversion distances at third positions to minimize effects of saturation and selection. The estimated age of *Potamotrygon* agrees with putative fossil evidence for this taxon¹². **b**, Distribution of *H. pacifica* and *H. schmardae* (black areas along coastlines), the closest marine relatives of the freshwater rays. The group is found in the Caribbean (the proposed source of Miocene marine incursions) but not the mouth of the Amazon (red arrow) or along the Pacific coast of South America (other proposed points of colonization).

freshwater stingrays does not occur along the Pacific coast of South America (Fig. 1b).

A diverse cross-section of the South American river fauna, including dolphins, fishes, crabs and snails, appear to be derived from marine ancestors. If these taxa originated contemporaneously with freshwater stingrays, the Miocene marine transgressions of South America will have had a profound effect on the diversification and structuring of neotropical communities.

Nathan R. Lovejoy*,

Eldredge Bermingham†, Andrew P. Martin‡

*Section of Ecology and Systematics,
Corson Hall, Cornell University,

Ithaca, New York 14853-2701, USA

†Smithsonian Tropical Research Institute,
Naos Marine Laboratory,

PO Box 2072, Balboa, Republic of Panama

‡Department of EPOB, CB 334,

University of Colorado,

Boulder, Colorado 80309, USA

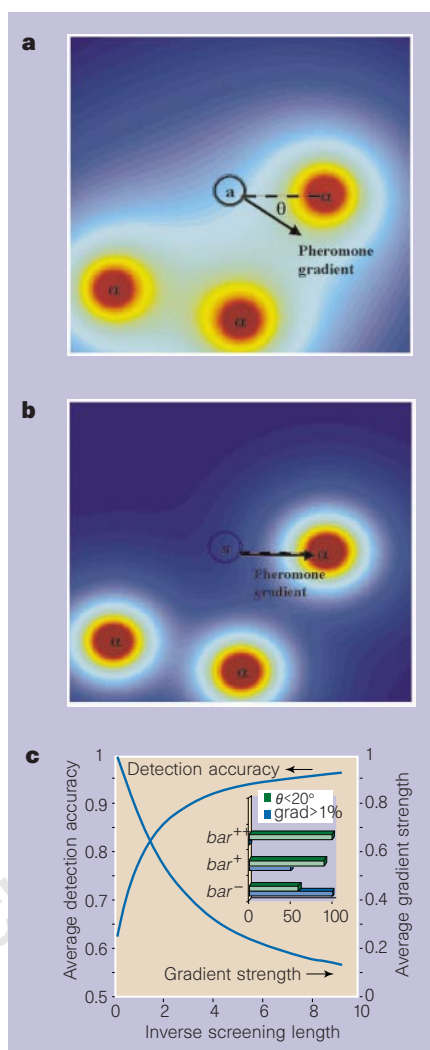
e-mail: am@stripe.colorado.edu

1. Roberts, T. R. *Bull. Mus. Comp. Zool.* **143**, 117–147 (1972).
2. Hoorn, C. *et al. Geology* **23**, 237–240 (1995).
3. Lundberg, J. G. *et al. in Phylogeny and Classification of Neotropical Fishes* (eds Malabarba, L. R., Reis, R. E., Vari, R. P., Lucena, C. A. S. & Lucena, Z. M. S.) (Museu de Ciências e Tecnologia, Porto Alegre, Brazil, in the press).
4. Nuttall, C. P. *Bull. Br. Mus. Nat. Hist.* **45**, 165–371 (1990).
5. Hoorn, C. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **105**, 267–309 (1993).
6. Croizat, L. *Space, Time, Form: The Biological Synthesis* (Published by the author, Caracas, Venezuela, 1962).
7. Brooks, D. R., Thorson, T. B. & Mayes, M. A. in *Advances in Cladistics* (eds Funk, V. A. & Brooks, D. R.) 147–175 (New York Botanical Garden, 1981).
8. Grabert, H. *Natur Mus.* **113**, 61–71 (1983).
9. Kroonenberg, S. B., Bakker, J. G. M. & van der Wiel, A. M. *Geol. Mijnb.* **69**, 279–290 (1990).
10. Swofford, D. L. *PAUP: Phylogenetic Analysis Using Parsimony, version 3.1* (Illinois Natural History Survey, Champaign, 1993).
11. Lovejoy, N. R. *Zool. J. Linn. Soc.* **117**, 207–257 (1996).
12. Lundberg, J. G. in *Phylogeny and Classification of Neotropical Fishes* (eds Malabarba, L. R., Reis, R. E., Vari, R. P., Lucena, C. A. S. & Lucena, Z. M. S.) (Museu de Ciências e Tecnologia, Porto Alegre, Brazil, in the press).

Protease helps yeast find mating partners

The choice of mating partner by the yeast *Saccharomyces cerevisiae* involves the detection of mating pheromones produced by other yeast cells. A cell that is capable of mating deduces the position of its nearest mating partner from the spatial gradient of pheromone. While studying this process, we realized that, in the presence of many potential mating partners, the gradient might not point in the direction of the nearest partner. Here we show that degradation of some of the mating pheromone by protease enzymes helps to align the gradient in the direction of the nearest partner, which increases mating efficiency.

Haploid yeast cells of the two mating types, *a* and α , signal to each other by the secretion of cell-type-specific pheromones



(*a*-factor and α -factor, respectively), which are detected by cells of the opposite mating type, triggering a series of events that lead to mating. In particular, the cell response includes directed growth towards the mating partner, the position of which is estimated from the direction of maximum pheromone concentration^{1–3}. Because yeast cells do not move, an appropriate choice of growth direction is essential for efficient mating.

Cells of mating type *a* secrete a protease that hydrolyses α -factor^{4,5}. Mutants deficient in this 'barrier activity' (*bar1* mutants^{6,7}) are highly sensitive to pheromones⁷, so the protease was thought to be involved in the recovery of yeast cells from the pheromone-induced cell-cycle arrest that is part of the pre-mating response. But *bar1* mutants also mate less efficiently with α -cells in a mass mating mixture¹. A role for the protease in mating-partner detection has also been suggested⁴.

To understand how the protease may increase mating efficiency, consider an *a*-cell surrounded by several α -cells, which secrete α -factor (Fig. 1a,b). These cells are part of a larger population of *a*- and α -cells. The steady-state profile of pheromone concentration between cells, $[P(\mathbf{r})]$, where $\mathbf{r}$

represents a position in three-dimensional space, is given by solving the diffusion equation $\partial[P(\mathbf{r})]/\partial t = D\nabla^2[P(\mathbf{r})] = 0$, where D is the diffusion coefficient. It is useful to consider an analogy with an electrostatic system, as the electrostatic potential satisfies the above equation. The α -cells can be thought of as electrostatic point charges, the pheromone concentration as an electrostatic potential, and the pheromone gradient as an electric field. The pheromone profile is given by: $[P(\mathbf{r})] = \sum_{i=1}^N \phi(\mathbf{r} - \mathbf{r}_i)$, $\phi(\mathbf{r}) = 1/|\mathbf{r}|$ where $\phi(\mathbf{r} - \mathbf{r}_i)$ is the contribution of the *i*th cell, located at $\mathbf{r}_i$, to the concentration. Because $1/r$ is a slowly decaying function of the radial distance, r , the pheromone concentration at any point is determined by the contribution of many α -cells. In general, therefore, the pheromone gradient at the location of an *a*-cell will not point in the direction of the nearest mating partner, but at an angle, θ , from it (Fig. 1a). But what is the role of the protease? Consider first-order degradation: $\partial[P(\mathbf{r})]/\partial t = D\nabla^2[P(\mathbf{r})] - k[B][P(\mathbf{r})] = 0$ where $[B]$ is the concentration of the *bar* protease, and k is a kinetic rate constant. In terms of the electrostatic analogy, the new

term, $-k[B][P(r)]$, corresponds to an induced screening charge in a conducting medium. For a uniform concentration of protease, the long-range $1/r$ potential is reduced exponentially:

$$\phi(r) = 1/r \rightarrow \phi_{scr}(r) = (1/r)\exp(-r/\lambda)$$

and $\lambda^2 = D/k[B]$.

In simple terms, degradation by the protease limits the size of the pheromone cloud around each α -cell to a radius λ , the average distance that a pheromone molecule diffuses before being degraded. This screening increases the alignment between the gradient and the direction to the nearest mating partner, although the gradient strength is reduced (Fig. 1b,c). Consequently, the protease is not helpful when there is only a single source of pheromone, as was observed by Segall³. Effective screening is obtained when λ is of the order of the mean distance between cells (about 10 μm), which for a diffusion-limited process indicates a protease concentration of about 1 nM.

For simplicity, we assumed uniform protease concentration when deriving the last equation. However, our results generalize to non-uniform conditions provided that the protease is present over a range corresponding to several intercell distances around the α -cell. This is likely to occur, as protease-secreting α -cells are distributed throughout the population and the protease is a stable, widely diffusing molecule^{4,8}.

Mechanisms that screen concentration gradients may be more generally applicable. Cells of mating type α may have an activity that inactivates α -factor⁹. Screening may also be used in chemotaxis; for example, the slime mould *Dictyostelium* secretes factors that antagonize the attractants cyclic AMP and folate. The effective range of signalling by diffusible growth factors may also be regulated by a similar mechanism. For example, the growth factor Spitz is inhibited in the developing *Drosophila* eye¹⁰. Studies of model systems, such as that discussed here, may provide a better understanding of the quantitative effects of mechanisms that control the range of signalling.

Naama Barkai*†, Mark D. Rose†, Ned S. Wingreen‡

*Departments of *Physics and †Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA*
 ‡NEC Research Institute, 4 Independence Way, Princeton, New Jersey 08540, USA
 e-mail: wingreen@research.nj.nec.com

1. Jackson, C. & Hartwell, L. H. *Cell* **63**, 1039–1051 (1990).
2. Jackson, C. & Hartwell, L. H. *Mol. Cell. Biol.* **10**, 787–796 (1990).
3. Segall, J. E. *Proc. Natl Acad. Sci. USA* **90**, 8332–8336 (1993).
4. Hicks, J. B. & Herskowitz, I. *Nature* **260**, 246–248 (1976).
5. Ciepek, E. & Thorne, J. *Cell* **18**, 623–635 (1979).
6. Sprague, G. F. & Herskowitz, I. *J. Mol. Biol.* **153**, 305–321 (1981).
7. Chan, R. K. & Otte, C. A. *Mol. Cell. Biol.* **2**, 11–20 (1982).
8. Manney, T. R. *J. Bacteriol.* **155**, 291–301 (1983).
9. Steden, M., Betz, R. & Duntze, W. *Mol. Gen. Genet.* **219**, 439–444 (1989).
10. Wasserman, D. & Freeman, M. *Trends Cell Biol.* **7**, 431–436 (1997).

Robust dinosaur phylogeny?

The Ankylosauria comprises two families of armoured dinosaurs (Nodosauridae and Ankylosauridae) that are best known from well-preserved specimens from the Cretaceous period. In their report on the skull of a new Jurassic ankylosaur, *Gargoyleosaurus*, Carpenter *et al.*¹ presented a phylogenetic analysis of four terminal taxa, which yielded a tree with *Gargoyleosaurus* as the sister taxon of the Ankylosauridae. But the authors' claim that their tree is robust is undermined when their data and their tree are evaluated using numerical techniques.

Although Carpenter *et al.*'s preferred tree (Fig. 1a) is the most parsimonious tree for their data, the alternatives in which *Gargoyleosaurus* is the sister taxon of either the Nodosauridae (Fig. 1b) or Nodosauridae + Ankylosauridae (Fig. 1c) are only slightly less parsimonious. Of their 26 characters, just six support their preferred tree, four support *Gargoyleosaurus* + Nodosauridae, and two support Nodosauridae and Ankylosauridae. The remaining 14 characters are phylogenetically uninformative.

The 'winning sites' test² shows that the differences between Carpenter *et al.*'s preferred tree and the slightly less parsimonious alternatives shown in Fig. 1b and c are not significant (*P* values of 0.75 and 0.29, respectively), so they do not allow us to identify any tree as being significantly better supported by the data than any other. Similarly, bootstrap support for the association of *Gargoyleosaurus* and the Ankylosauridae is not compelling (71%, 2,000 replicates). Matrix randomization tests also indicate the limitations of the data. Permutation tail probabilities for the data based on parsimony tree length³ and pairwise character compatibilities^{4,5} are 0.51 and 0.43, respectively, and do not allow us to reject the null hypothesis that congruence within the data is no greater than expected by chance alone.

Taken together, these results indicate that Carpenter *et al.*'s claim that their preferred tree is robust is not justified. Their conclusions may also be undermined by their use of aggregate in-group terminal taxa (nodosaurids and ankylosaurids), which precludes the placement of *Gargoyleosaurus* within either of these clades. Consequently, a phylogenetically important position for *Gargoyleosaurus* as sister taxon to a major clade is inevitable.

Choice of taxa also has an effect on the character evidence. The assumed plesiomorphic condition of character 22 (a 'neck' at the base of the occipital condyle),

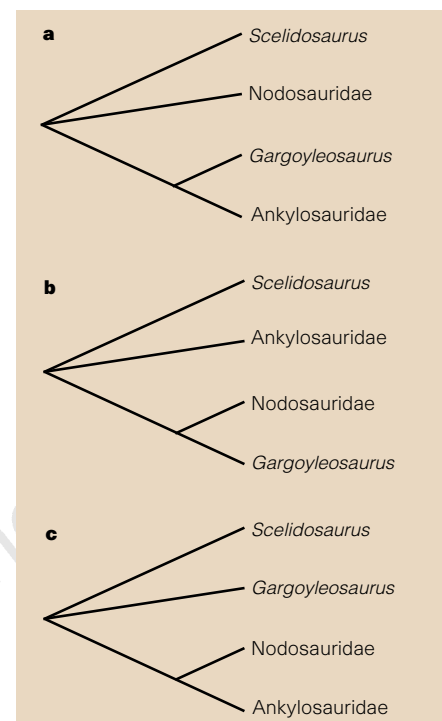


Figure 1 The three possible trees of ankylosaur relationships (rooted on *Scelidosaurus* with the basal polytomy unresolved). **a**, Tree based on the data of Carpenter *et al.*¹, which is the most parsimonious tree (tree length of 32 steps); **b**, the next most parsimonious tree (tree length of 34 steps); **c**, the least parsimonious tree (tree length of 36 steps).

one of the six characters used to support *Gargoyleosaurus* + Ankylosauridae, is present in several ankylosaurids (for example, *Talarurus*, *Tarchia* and *Maleevus*)⁶. Recording this character as polymorphic for the aggregate Ankylosauridae reduces the number of potential synapomorphies for *Gargoyleosaurus* + Ankylosauridae to just five, and highlights the possibility that *Gargoyleosaurus* might nest within the Ankylosauridae.

Carpenter *et al.*¹ noted that strong brain flexure in *Gargoyleosaurus* is shared with at least some nodosaurids, but they interpreted this similarity as primitive because the ornithomimid *Hypsilophodon* has a similar condition. In contrast, comparison with Stegosauria, a more proximate outgroup, suggests that this similarity is derived within Ankylosauria, increasing the evidence for *Gargoyleosaurus* + Nodosauridae to five potential synapomorphies. That these minor revisions to the character data can render the hypotheses represented in Fig. 1a and b equally parsimonious underscores the frailty of the data and the phylogenetic conclusions.

Gargoyleosaurus may well be an ankylosaurid, but this inference has not yet been demonstrated. Robust resolution of the phylogenetic placement of *Gargoyleosaurus* and assessment of its implications for the evolution of Ankylosauria will require a

thorough analysis of a more comprehensive data set.

Mark Wilkinson[†], Paul Upchurch^{*},
Paul M. Barrett[‡], David J. Gower[§],
Michael J. Benton[§]

^{*}School of Biological Sciences,
University of Bristol,
Bristol BS8 1UG, UK

[†]Department of Zoology,
Natural History Museum,
London SW7 5BD, UK

[‡]Department of Earth Sciences,
University of Cambridge,
Cambridge CB2 3EQ, UK

[§]Department of Earth Sciences,
University of Bristol,
Bristol BS8 1RJ, UK

1. Carpenter, K., Miles, C. & Cloward, K. *Nature* **393**, 782–783 (1998).
2. Prager, E. M. & Wilson, A. C. *J. Mol. Evol.* **27**, 326–325 (1988).
3. Faith, D. P. & Cranston, P. S. *Cladistics* **7**, 1–28 (1991).
4. Alroy, J. *Syst. Biol.* **43**, 430–437 (1994).
5. Wilkinson, M. *Biol. Rev.* **72**, 423–470 (1997).
6. Barrett, P. M., Hailu, Y., Upchurch, P. & Burton, A. C. *J. Vert. Paleontol.* **18**, 376–384 (1998).

Moving ahead through differential visual latency

The time it takes to transmit information along the human visual pathways introduces a substantial delay in the processing of images that fall on the retina. This visual latency might be expected to cause a moving object to be perceived at a position behind its actual one, disrupting the accuracy of visually guided motor actions such as catching or hitting, but this does not happen. It has been proposed that the perceived position of a moving object is extrapolated forwards in time to compensate for the delay in visual processing^{1–3}.

We have studied the spatial misalignment perceived between moving and strobed objects and find that it varies systematically with the luminance of the objects. Our results favour an explanation for these perceived misalignments based on differential visual latencies, rather than on motion extrapolation. Thus, accurate visually guided motor actions are likely to depend on motor instead of perceptual compensation.

Evidence for a mechanism based on motion extrapolation^{1–3} comes from the flash-lag phenomenon⁴, in which a continuously moving object is perceived to be ahead of a stationary strobed object when the two retinal images are physically aligned. But because visual latency varies according to the properties of a stimulus, including its luminance^{5–9}, this mechanism would have to compensate appropriately for a range of stimulus-dependent variations in latency to ensure that real-time, visually guided responses are accurate. An alternative, previous

explanation, invoking a longer delay for the processing of a flashing stimulus, was based on attentional mechanisms¹⁰.

According to the hypothesis based on differential visual latencies, the observed spatial lead of the moving central segment in Fig. 1a is directly proportional to the difference between the latencies of the strobed and the moving central segments. For a given stimulus, the visual latency varies inversely with its luminance^{5–9}, so the observed spatial lead in the flash-lag paradigm should vary according to the luminance of the strobed and moving central segments. Increasing the luminance of the moving central segments but not that of the strobed segments should decrease the delay of the moving central segment (d_m) while

that of the strobed segment (d_s) remains constant. The latency-difference hypothesis therefore predicts that the observed spatial lead of the moving central segment should increase.

To test this prediction, we measured the spatial lead of the moving central segment as a function of the detectability of the central segment while keeping the detectability of the strobed segments constant. Here we use detectability to refer to the number of log units of luminance (Lu) above the detection threshold; detectability of the strobed segments was 0.3 Lu for subjects S.S.P. and G.P., and 0.5 Lu for T.L.N. The temporal lead of the moving central segment averaged across subjects increases systematically from 20 to 70 ms when its detectability increases by 1.0 Lu (Fig. 1b).

Increasing the luminance of the strobed segments while keeping that of the moving central segment constant should decrease d_s , while d_m remains constant. The latency-difference hypothesis predicts that the observed spatial lead of the moving central segment should decrease and, if the luminance of the strobed segments is high enough, the moving central segment should be perceived to lag behind spatially. We tested this prediction by measuring spatial lead as a function of the detectability of the strobed segments, while keeping the detectability of the moving central segment constant (1.5 Lu above the detection threshold for subjects G.P. and T.L.N., and 0.8 Lu for S.S.P.). The observed temporal lead of the moving central segment averaged across subjects decreases systematically from 80 to –30 ms as the detectability of the strobed segments increases by 1.5 to 2.0 Lu (Fig. 1c).

These results support predictions of the latency-difference hypothesis and show that the motion-extrapolation mechanism does not compensate for stimulus-dependent variations in latency. Indeed, theoretical calculations show that the putative motion-extrapolation mechanism must be under-compensating by at least 120 ms to account for the data in Fig. 1. But a motion-extrapolation mechanism that does not adequately compensate for variations in visual latency would not appreciably improve the accuracy of real-time visually guided behaviour.

Gopathy Purushothaman^{*}, Saumil S. Patel[†],
Harold E. Bedell[†], Haluk Ogmen^{*}

^{*}College of Engineering and [†]College of Optometry,
University of Houston,
Houston, Texas 77204, USA
e-mail: hbedell@uh.edu

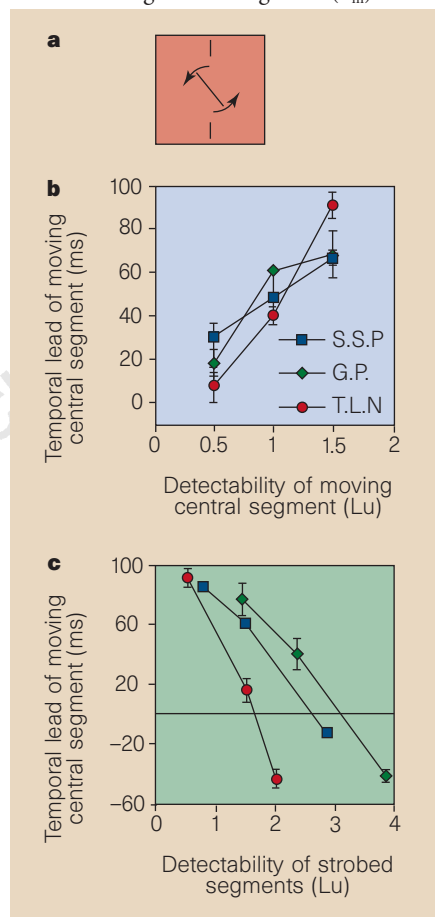


Figure 1 Luminance-dependent misalignments between moving and strobed targets. **a**, The stimulus was a continuously rotating central segment (40 rev min⁻¹) and two flanking strobed segments (5 ms). **b**, The observed temporal lead of the moving central segment is shown as a function of its detectability (0 = threshold) for three subjects (S.S.P., G.P. and T.L.N.). Positive values on the y-axis represent a temporal lead in perceiving the moving central segment relative to the strobed segments. The observed spatial lead was converted into a temporal lead by dividing it by the velocity of the moving central segment. **c**, The observed temporal lead of the moving central segment is shown as a function of the detectability of the strobed segments (0 = threshold) for the same subjects.

1. Nijhawan, R. *Nature* **370**, 256–257 (1994).
2. Nijhawan, R. *Nature* **386**, 66–69 (1997).
3. Khurana, B. & Nijhawan, R. *Nature* **378**, 566 (1995).
4. MacKay, D. M. *Nature* **181**, 507–508 (1958).
5. Bacon, F. *Novum Organum* (Bills, London, 1620).
6. Cattell, J. M. *Brain* **9**, 512–515 (1886).
7. Mansfield, R. J. W. *Vis. Res.* **13**, 2219–2234 (1973).
8. Roufs, J. A. J. *Vis. Res.* **14**, 853–869 (1974).
9. Williams, J. M. & Lit, A. *Vis. Res.* **23**, 171–179 (1983).
10. Baldo, M. V. C. & Klein, S. A. *Nature* **378**, 565–566 (1995).

Assimilating shocks to the system

Frankenstein's Children: Electricity, Exhibition, and Experiment in Early-Nineteenth-Century London

by Iwan Rhys Morus

Princeton University Press: 1998. 324 pp. \$45, £32.50

Patricia Fara

Around the end of the eighteenth century, propagandizing authors started to tell success stories about the new scientific disciplines, attributing their rise to key discoveries made by individual founding fathers. The commemoration of these scientific heroes was tinged with nationalism: the statue of Isaac Newton erected in Grantham in 1858 was made from melted-down cannon captured during the Crimean War, thus literally embodying England's military prowess as well as advertising her intellectual supremacy. Ten years later, the government cited Newton's statue, paid for by private subscription, as a precedent justifying its refusal to fund a monument to Michael Faraday. But Faraday's elite supporters at the Royal Institution, keen to immortalize him as a great Victorian scientist, financed a statue showing this autodidactic blacksmith's son swathed in an Oxford academic gown, an electromagnetic coil in his hand.

But the past is constantly open to reinterpretation, and less heroic narratives of science's history can also be created. In his impressive, revisionary *Frankenstein's Children*, Iwan Rhys Morus convincingly undercuts traditional versions of early nineteenth-century electricity in several important respects. Instead of perpetuating the myth of Faraday as a working-class genius, Morus exposes his image-fashioning techniques and places him in a broader context of skilled contemporaries who were displaying the marvels of electricity to less distinguished audiences. By refusing to take for granted such dichotomies as science/technology or discovery/invention, Morus reveals how science's history is also the history of our industrialized society. Hans Oersted famously made his compass needle flicker on a laboratory bench, but converting that romanticized experiment into pervasive and indispensable electromagnetic equipment involved marshalling human as well as material and financial resources.

Still more ambitiously, Morus has tackled what is often called 'the big picture' problem. During the past thirty years or so, historians and sociologists have reacted against traditional accounts of science's progressive march towards absolute truth by trying to deprive and demystify the processes through which scientific results come to be believed. They have painstakingly analysed



Switched on to a cure: Medical therapists galvanized their patients back to health.

specific episodes in minute detail to demonstrate the multiplicity of cultural, political and economic influences that affect scientists and their interactions with society. While this approach has yielded invaluable insights into scientific practices, these microhistories need to be integrated into macrohistorical overviews that examine, but do not necessarily celebrate, how science has come to acquire such a prestigious status during the past couple of centuries.

Approaching his topic from two complementary perspectives, Morus provides a colourful but subtle account of electricity's cultural and economic importance in early nineteenth-century London. As he wryly observes, it is hard to undermine heroic histories of electricity while focusing on Faraday's predominating presence. Seeking to escape from this biographical straitjacket, he devotes the first half of his book to exploring the places where electrical experiments were located, although he does still write extensively about who performed them. Faraday took elocution lessons so that he could better purvey packaged knowledge to the leisured middle classes flocking to his Friday evening lectures at the Royal Institution (until the floor started caving in under their weight). His critical rival William Sturgeon came from a similar background, but cultivated less elite audiences. Now largely forgotten, Sturgeon was highly respected by many eminent contemporaries for his demonstrative apparatus which deliberately revealed its internal operations and was to prove crucial for the development of electrical technology

in the second half of the century.

Ensconced in his private basement, Faraday devised displays designed to dazzle wealthy spectators with his philosophical virtuosity, but Sturgeon and many other popular lecturers provided entertaining instruction in less prestigious theatres. More radical practitioners believed that electrical knowledge should be freely available and dedicated to utilitarian ends. Enticing audiences with electromagnetic models of the Earth or machines that fired guns, decomposed water and shocked visitors' tongues, skilled artisans supported themselves through their electrical inventions displayed in these public laboratories. Medical therapists galvanized their patients back into health, even — echoing *Frankenstein* — proffering the hope of restoring the dead to life.

This emphasis on spectacle also permeates the second half of Morus's book, in which he explores how electricity emerged from these diverse exhibition spaces to enter the commercial world of early nineteenth-century England. Entrepreneurial experimenters struggling to earn a living fashioned themselves as respectable inventors and carved out new careers in a society that was becoming increasingly dependent on technology. Flamboyant salesmen lit up Nelson's statue in Trafalgar Square, advertised long-distance chess games played over the new telegraph system, and marketed galvanic machines and jewellery. These promotional gimmicks helped less flashy electrical devices to become incorporated in the machine cul-

ture of the Victorian age. As the rapidly strengthening patent system protected innovators from rapacious competitors, new industries sprang up that provided mass-produced, electro-plated domestic products, while telegraph networks revolutionized not only private and business communications, but also functioned like a giant global nervous system, binding together the flourishing Empire.

The outcome of many years of archival research, and benefiting from sources that were formerly neglected, such as popular journals, *Frankenstein's Children* is analytically extremely sophisticated, packed with details, yet eminently readable. This fascinating illustrated account of early Victorian culture rescues the lives and activities of artisans whose traces have been obliterated from our collective scientific memory. At the same time, as Morus could perhaps have drawn out more explicitly, his historical study of the social negotiations that underpinned technological innovation in the past also illuminates the contingent nature of so many aspects of the modern, consumer-based electronic world that we take for granted. □

Patricia Fara is at the Max-Planck-Institut für Wissenschaftsgeschichte, Wilhelmstrasse 44, 10117 Berlin, Germany.

Charting the stormy seas of AI

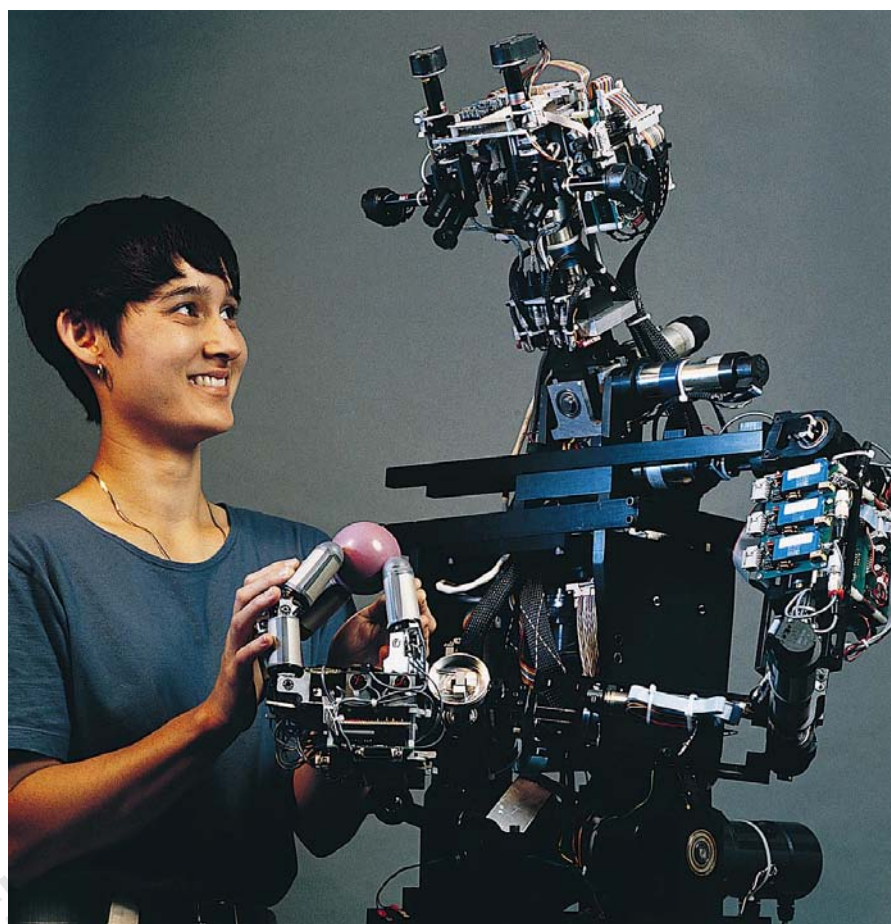
Mapping Great Debates: Can Computers Think?

project led by Robert E. Horn
Mapping Great Debates Project. MacroVU Press: 1998. Set of 7 maps and a handbook. \$99, £60 (introductory price) \$125, £75 (from January 1999)

Harry Collins

The long discussion about whether electronic computers can think, begun by Alan Turing, is represented here on seven charts or 'maps'. Each map is about 4 feet wide and 3 feet high, so you need a large wall such as a corridor in a school or university to see the whole thing. The maps have to be displayed at eye level because each sheet reveals around 130 inter-linked boxes, each box representing a position in an argument, and the writing in the boxes is about the size of the print on this page. Key arguments are supported by photographs of their authors, or other illustrations or cartoons. Turing's picture appears at the top left-hand corner of every sheet and I was thrilled to find my mug staring out towards the bottom right of sheet one — so I'm positively disposed towards the enterprise.

I also like the idea because of the discipline it forces on the editors. Each box must present an argument in a few words and each must be linked into a network of argument,



Keep in touch: The android Cog is an artificial intelligence learning to manipulate a body.

ramifying forwards, backwards and sideways. This leaves no room for mush; the steps have to be exact — more like the lines in a computer program than the discursive form of ordinary text. Some people use the flexibility of ordinary text as a way of avoiding the necessity of working out exactly what they want to say; that sort of thing can't pass muster here.

There are, however, downsides to this exactitude. The fundamental units that are assembled to make the charts are bite-sized contributions to strands of debate rather than developed 'positions'. I think of the AI debate as having historical roots with a few strong branches springing from Turing's trunk, each branch represented by a major book, paper or writer. For the charts, the tree has been cut down, splintered and glued together again, like chipboard, and the historical role of the players is lost to view. Thus, for me, Hubert Dreyfus wrote the earliest decisive critique of symbolic AI and is the towering figure on the 'anti' side, but here you find him represented quite late in the sequence of maps by discrete elements of his argument.

Strange juxtapositions also arise. In the course of developing a position such as 'computers cannot think', X will sometimes criticize Y's similar argument because X thinks his own argument is better. If X's criticism of

Y's argument is novel or unique it might well appear on the chart in a place that makes X look like a supporter of the thinking power of computers even though he is well known for the opposite view. This is not a fatal flaw, but it is a curiosity.

Another disadvantage is that the treatment favours 'bitty' debates. Thus, the argument about whether the Turing test is valid occupies a whole sheet while Searle's Chinese Room, which is just a variant of it, occupies another whole sheet. This amount of space is disproportionate to the intellectual significance of these discussions; they are well known but they are not deep currents of thought. They do, however, invite many small contributions, for which this medium is a natural fit. More subtle currents, such as that concerned with the socially situated nature of human thinking and action, are less easy to represent in this format, and we lose sight of the way they engage with those great swathes of bitty discussion.

Another reason for concern is that there is something more authoritative about a 'map' than a text — it seems more representative of a fixed reality. Furthermore, the charts will be displayed in public and those who do not want to read more deeply will use them as a 'crib'. And, unlike texts, they will have no competitors putting forward alternative views, especially at \$125. For these

reasons the editing in such an enterprise is very important.

The honesty of the editing comes through in the project's very refusal to stick to the usual diet of the great and the good. One must also appreciate the sheer amount of work that has gone into the preparation. But I am pleased to see that a second edition is planned and that readers are invited to write in with their comments and contributions. If this kind of enterprise is to deserve the kind of authority it might well gain, the second edition, and each of the future exercises which are at the planning stage, should have a prominently named editorial panel of experts who are seen to be taking responsibility for what the charts say. Debates such as these are passionate affairs conducted by humans arguing from within quite different intellectual contexts; represent them how you like, they are not computer programs — but then, as a sociologist, I would say that, wouldn't I? □

Harry Collins is at the Centre for the Study of Knowledge Expertise and Science, Cardiff University, 50 Park Place, Cardiff CF1 3AT, UK.

Heisenberg revisited

Heisenberg and the Nazi Atomic Bomb Project: A study in German Culture

by Paul Lawrence Rose
University of California Press: 1998. 351 pp.
\$35, £25

Mark Walker

The controversy surrounding Werner Heisenberg, nuclear weapons and National Socialism appears unending. *Heisenberg and the Nazi Atomic Bomb Project* is the latest in a long line of books on the subject, and more are yet to appear. Despite the proliferation of this literature, with only a few exceptions it can all be divided into two camps.

The 'sabotage' camp argues, to varying degrees, that Heisenberg and some of his colleagues actively resisted Hitler by somehow denying him nuclear weapons. This view was first put forward by Robert Jungk in *Brighter Than a Thousand Suns* (Harcourt Brace, 1958), and the most recent proponent is Thomas Powers in *Heisenberg's War* (Knopf, 1993). The 'incompetence' camp, on the other hand, argues that Heisenberg and colleagues made significant mistakes in their efforts to harness nuclear fission, and either implies or asserts that the claim of Heisenberg's resistance is thereby disproved. Samuel Goudsmit began the argument in *Alsos* (Henry Schumann, 1947) and Paul Rose is now the most vociferous proponent of the thesis.

Rose's book contains a historiographic survey of the literature on Heisenberg and the bomb; an examination of Heisenberg's

scientific work on the bomb; and a chapter on the German mentality in general, and Heisenberg's in particular. The historiographic survey is thorough and generally accurate and fair. But Rose does some of his secondary sources a disservice by emphasizing his disagreement with them in often strident language, while downplaying, if not omitting mention of, the considerable degree to which he not only agrees with them, but actually depends on them.

Rose's book is essentially a reinterpretation of existing literature. There is little new primary material, and even most of Rose's arguments can be found in other authors' work. Rose's original contribution is the intensity with which he pursues his goals.

Rose's presentation of the scientific evidence is more problematic, essentially because his approach is teleological. He sets out to prove that Heisenberg made significant scientific errors with regard to nuclear weapons. Anything that might support this interpretation is emphasized, while evidence that does not fit is explained away. But what is the point of this exercise? Such errors by Heisenberg would be important if they were the reason why Hitler never got nuclear weapons, but Rose freely admits that this was not the case. They might also be important if they shed light on Heisenberg's scientific legacy, on his methodology, or on any of the many areas of theoretical physics he worked in during his long and productive career, but Rose does not even consider this question.

Instead, Rose's investigation appears to be a continuation of the attacks Goudsmit made on Heisenberg's scientific reputation during their tumultuous postwar dialogue on the relationship between science and the

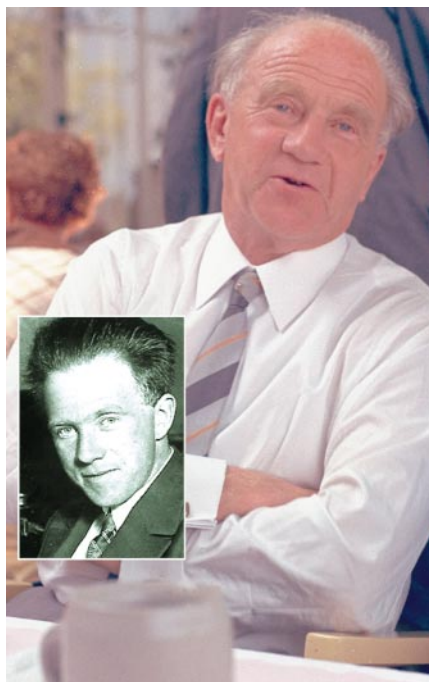
National Socialist state. Heisenberg's service for the Third Reich and, in particular, his postwar refusal to admit that he had done anything wrong, alienated Goudsmit and a series of scientists and historians who took up Goudsmit's cause long after both men were dead.

Rose does address the evidence that appears to contradict his thesis. His argument that Heisenberg made a significant error boils down to the following: although Heisenberg understood how a plutonium bomb would work in theory, he failed to understand how a device using uranium-235 as nuclear fuel would operate. This argument is inherently implausible and unconvincing, but readers can judge the matter for themselves. To see if Rose's argument is persuasive, read the translated text of the talk Heisenberg gave in February 1942 before representatives of the National Socialist party, the German government and industry, and look at an important diagram from a 1943 talk, in *Physics and National Socialism*, edited by Klaus Hentschel (Birkhäuser, 1996, pages 294–301).

Rose's third section begins with Martin Luther's ruthlessness and anti-Semitism and uses other historical examples to build the case that there is a peculiar German mentality, which is difficult for the Western mind to understand, and which predisposed Germans to serve National Socialism, one way or another. Here Rose's argument is reminiscent of Daniel Goldhagen's *Hitler's Willing Executioners* (Knopf, 1996), which argues that the Holocaust occurred because Germans hated the Jews. Rose's thesis predates Goldhagen's book, however, and is disturbing because it appears to be a mirror image of the National Socialists' own claims of a peculiar Jewish mentality. If anti-Semitism is bad, is Germanophobia better?

Rose carefully pursues every step of Heisenberg's often ignoble path through the Third Reich, emphasizing wherever possible events that could discredit him. Although Rose's analysis is persuasive, for Heisenberg did enter into a Faustian pact with Hitler's regime and, after the war, steadfastly rejected any moral criticism of his actions, the zeal with which Rose pursues Heisenberg is nevertheless striking. There were probably very few prominent German scientists during the Third Reich who could undergo such an investigation without having their reputations tarnished.

Rose's book does not really have a conclusion, and it is not clear that he could have written a suitable one. When all the detail of his analysis and narrative is processed, and all his arguments weighed, they boil down to this: Heisenberg was no hero, certainly no resistance fighter, and after the war he never owned up to his moral and, in Rose's view, scientific shortcomings. But is this really a revelation, surprising or profound?



Heisenberg: No hero in the war years (inset) and no comment in peace time (main pic).

The true shortcoming of this book is that, despite Rose's attempts to penetrate Heisenberg's "German mentality", the author's prosecutorial analysis gives the reader little understanding of Heisenberg as a human being, or of how difficult it was to live and work under such a regime. □

Mark Walker is in the Department of History, Union College, Schenectady, New York, 12308-3163, USA

The science of patterns?

The Language of Mathematics: Making the Invisible Visible

by Keith Devlin

W. H. Freeman: 1998. 344 pp. \$24.95, £14.95

Jeremy Gray

In recent years there has been a profusion of books aimed at making mathematics comprehensible and interesting to the general reader. Keith Devlin is the successful author of *Mathematics: The Science of Patterns* and this book is a spin-off from that one, aimed at

a larger audience. Devlin has taken the opportunity to add two new chapters, one on chance and one on space-time, but has had to manage without the profusion of artwork and colour photographs that adorned the earlier book.

He covers a wide range of topics with an enviably light touch. The fact that there are infinitely many prime numbers sits alongside tests to see if a number is prime, and leads into a discussion of public key encryption systems (unhappily the details of the system are not explained). There is a chapter on logic and set theory, topics on which Devlin is an expert, and this leads to a hint or two about Noam Chomsky's ideas on linguistic structures. There is some calculus and Newtonian mechanics, some geometry, including Desargues's theorem and the idea of spaces in any number of dimensions.

We get some discussion of groups, including wallpaper patterns and lattices, which leads to the unresolved question of the densest packing of spheres. Topology leads to the theory of knots, the Jones polynomial which almost classifies them, and their relation to the study of DNA. Fermat's Last Theorem passes by, as does a discussion of



Point of infinity: Renaissance artists discovered how to create a sense of depth in a painting.

chance, from Quetelet's 'average man' to the Black-Scholes formula much used in the derivatives market. The book closes with a quick trip through gravity, electromagnetic theory, and the idea of space-time.

All this is done with clarity and wit, a novel collection of puns, and a willingness to cut corners. The history of mathematics is taken up when it helps and dropped when it does not; it is accurate when it is easy and less so when simplification is judged appropriate to the greater aim. The same is true of the mathematics, and the subjects to which it is applied. The book is weakest when discussing these applications. Devlin sometimes stops short, for fear of becoming too hard, before he has really become comprehensible. Inevitably, some of the topics are familiar in the genre, while others are new, but even the old ones are deservedly popular. Even mathematics has only so many greatest hits. But it would be a churlish reader who did not find that mathematics is much more diverse than might be suspected, and much more useful in all sorts of ways.

As with his earlier book, Devlin pushes the quick definition that mathematics is the science of patterns. Readers may not be persuaded, but they should certainly come away thinking that mathematicians have something attractive to say about their subject, and about many others besides. Sadly, they will not find notes and a bibliography documenting where all this information came from, nor are they provided with suggestions for further reading. □

Jeremy Gray is in the Faculty of Mathematics and Computing, Open University, Milton Keynes MK7 6AA, UK.

New in paperback

T. rex and the Crater of Doom

by Walter Alvarez

Penguin, £7.99

"Alvarez portrays a modern geology that builds on the interdisciplinary resources of physics, chemistry and astronomy to fashion a synthesis about how our planet has evolved. . . . One can read *T. rex and the Crater of Doom* in a single sitting and I recommend it highly — if only as a jumping-off point to other perspectives on this dramatic scientific revolution", Clark R. Chapman, *Nature* 387, 33 (1997).

Born To Rebel: Birth Order, Family Dynamics and Creative Lives

by Frank J. Sulloway

Abacus, £9.99

"*Born To Rebel* contains a treasure-trove of information. . . . I doubt that anyone will ever again claim that the variable of birth order has been shown to be unimportant", Howard Gardner, *Nature* 384, 125 (1996).

The Ovary of Eve: Egg and Sperm and Preformation

by Clara Pinto-Correia

University of Chicago Press, \$17.50, £13.95

"Her engrossing book is likely to remain for many years the most readable and informative — certainly the most enjoyable — account of the early scientific efforts to explain human conception. . . . It is a strength of *The Ovary of Eve* that it succeeds in fully appropriating the view-point of the scientists of the past, and makes the reader realize the nobility, the

earnestness, the inevitability and the beauty of many of their constructs, even those condescendingly dismissed as 'foolishness' by less perceptive historians", Frank Gonzalez-Crussi, *Nature* 390, 41 (1997).

Women Scientists in America, Volume Two: Before Affirmative Action, 1940-1972

by Margaret W. Rossiter

John Hopkins University Press, £15

"... assembled with the careful scholarship that has become Rossiter's hallmark. Once again, the quantity of material researched is enormous and my caveats are few. . . . I found the book heavier going than its predecessor, with fewer of the light touches of irony and humour that I had enjoyed so much before", June Goodfield, *Nature* 380, 306 (1996).

Imagined Worlds

by Freeman Dyson

Harvard University Press, \$14, £8.50

"Dyson has a startlingly profound imagination, a willingness to take ideas as far as they can possibly go. . . . I think Dyson could write a better book about the future than this one. I think he could write a better book about the future than almost anyone; I wish he would", Oliver Morton, *Nature* 387, 361 (1997).

Charles Darwin's Letters: A Selection 1825-1859

edited by Frederick Burkhardt

Cambridge University Press, £9.95, \$12.95

The vertebrate track record

Martin G. Lockley

A renaissance in the study of fossil footprints has been driven by a multitude of discoveries and the realization that vertebrate ichnology makes important contributions to our understanding of terrestrial vertebrates. More striking is the insight the track record gives us into bias and incompleteness in the vertebrate fossil record.

We recognize that the fossil record, especially that of terrestrial vertebrates, is incomplete, but is it as incomplete as we assume? A huge tetrapod track record fills many gaps in our knowledge of the distribution of fossil vertebrates¹⁻⁴, and research into tetrapod ichnology has contributed to a range of palaeontological subdisciplines. Ichnology has generated its own extensive database, which complements the skeletal record in some cases but offers new insights in others. The data point to biases and gaps in the tetrapod fossil record^{4,5}.

Defining and redefining the track record

What do tracks tell us about the fossil record? To answer this question effectively, we must redefine our understanding of the track record, evaluate what we have learned, and identify problems that still lie ahead. In the wake of the dinosaur renaissance⁶, the field of vertebrate ichnology has grown rapidly. Great progress has been made in documentation of numerous track sites on all continents¹⁻⁴ and in understanding tracks as essential for correct reconstructions of posture and gait^{7,8}. Ichnologists have also asked new questions about the usefulness of tracks for palaeoecological and biostratigraphical research, and shown that tracks help us to understand and measure bias in the body fossil record⁵.

Despite commanding attention early in the nineteenth century^{9,10}, vertebrate ichnology failed to mature until recently, leading to the perception that tracks are rare and of limited utility. As recently as 1962, only about 27 discrete dinosaur footprint localities were known worldwide¹¹. Now the number is more than 300 for the western USA⁴, with comparable numbers from elsewhere in the world^{1,10,12-16}. Fossil footprints are found abundantly in strata from the Carboniferous period to the Holocene epoch, and occur in almost all terrestrial sedimentary deposits⁴. Consequently, tetrapod ichnology has the potential to yield large samples^{5,15}. It is well known that many Palaeozoic red beds are track-rich and bone-poor¹⁷. The Mesozoic stages of the western USA provide an even better example of the magnitude of track samples. There are at least three 'dinosaur freeways' from the Jurassic and Cretaceous stages of this region^{3,4}. These are regionally extensive surfaces, or thin stratigraphic units, of area 1,000–10,000 km². Dinosaur track densities average about 1 million per km² (1 per m²) and densities in excess of 100 tracks per m² have been reported for birds and other small tetrapods. In the temporal dimension, many track-bearing layers occur with a frequency as high as two per metre in sections measuring hundreds of metres in thickness^{3,4,15}.

The Jurassic System of the western USA is a good example of such track abundance. More than 200 track sites have been reported from seven well-known stratigraphic units (Fig. 1). Three of these units contain no skeletal remains and two have only a very sparse skeletal fauna. Allowing for an average of ten individual trackways per track site, our total sample is about 2,000. Even after extracting non-dinosaurian footprints, such a figure compares favourably with the total of generically determined dinosaur individuals (about 2,000) recovered from the entire Mesozoic stage worldwide^{18,19}.

Tracks are traditionally seen as useful for demonstrating vertebrate presence and activity where skeletal remains are absent. This bone-centred view is inherited from students of body fossils, looking at ichnology as a possible means of addressing specific

questions unanswered by the skeletal evidence. Now that vertebrate ichnology has matured it poses its own empirically derived questions²⁰. From a track-centred viewpoint we appreciate better²⁰ that, in most cases, footprints are fundamentally different from skeletal evidence (Box 1).

Preservation problems and progress

An obstacle to the use of some tracks and trackways is their incompleteness, which may result from several factors. True tracks (those made on the actual surface available for ichnological study) may be incomplete if the substrate was too soft, too firm or too variable. Track depth can determine the number of digit impressions preserved¹⁻³: thus, tracks may include fewer digits than the maximum number preserved under optimal conditions. Track-makers may also overprint their front footprints with their hind footprints, giving rise to the concept of primary, secondary or tertiary overlap²¹ and the potential for quadrupeds to appear bipedal²².

Some footprints are transmitted into underlayers as underprints or ghost prints^{2,3}, which may not reflect the precise anatomy of the trackmaker's foot. Experiments duplicate and describe accurately the range of morphological variation found in the subfossil record and show the relationship of variation to substrate type²³.

Despite views to the contrary²⁴, we are now reaching a consensus that ichnotaxonomic names should be assigned only to well-preserved tracks that reflect the morphology of the trackmaker^{17,18}. This does not, however, imply that we must know the identity of the trackmaker.

Careful description of the three-dimensional morphology and sedimentological context of tracks should be encouraged, to establish the palaeoenvironmental setting of tracks. Sedimentologists have yet to take full advantage of what vertebrate ichnology can teach about the palaeoenvironment. Exceptions are Nadon and Issler²⁵, who use dinosaur tracks as 'palaeopenetrometers' (tools for measuring compaction in floodplain sediments).

What can be learned from tracks?

Studies of locomotion based on tracks precipitated a revival of tetrapod ichnology with a debate over the speeds attained by dinosaurs²⁶. Estimates of rapid progression of large dinosaurs

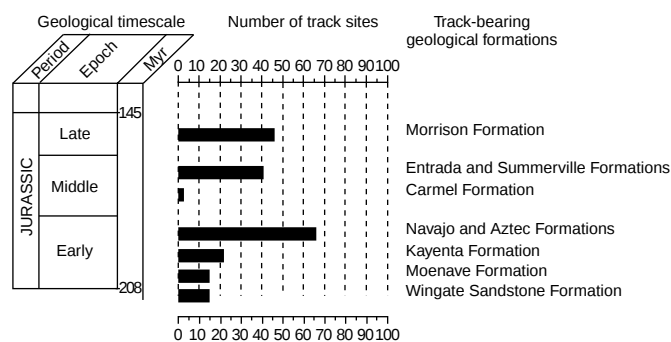
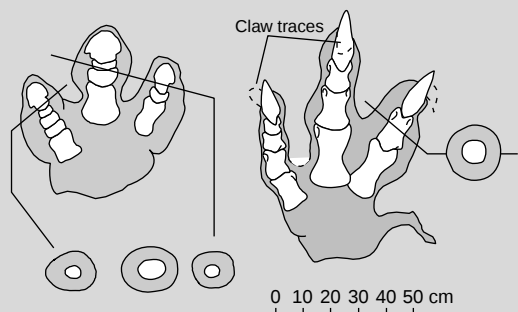


Figure 1 About 200 Jurassic track sites have been recorded in a small part of the western United States (mainly eastern Utah and Colorado), most during the last decade. At an estimated ten trackways per site, the total trackway sample is ~2,000 individuals.

Box 1 The anatomy and naming of tracks

Do footprints mirror the morphology of foot skeletons? Do foot skeletons even exist to match footprints? Consider the example of a probable *Tyrannosaurus rex* print. At 85 cm in length, this track is the largest theropod track known⁶³. It reveals a fleshy, well-padded foot, as expected for a 6- or 7-tonne animal, but the foot skeleton (digits II–IV) does not fit the outline of the track perfectly; it occupies only a small part of the cross-sectional area of the foot, as is the case in many large animals (see figure). For example, the reconstructed foot skeleton of a rhinoceros⁶⁴ or elephant indicates a digitigrade or unguligrade animal⁶⁴ but is of little use in predicting footprint morphology without the large fleshy portion of the foot being available. In general, the larger the trackmaker, the fleshier the foot, and the greater the morphological difference between the footprint and the foot skeleton. This observation is fundamental in justifying the use of ichnotaxonomy for naming well-preserved footprint morphologies. Where footprints are known without corresponding body fossil remains, as was the case for the 'hand animal' *Chirotherium* from 1835–1965 (ref. 52), the use of ichnotaxonomy is crucial.

The conclusion that tracks are different from foot skeletons might suggest limited potential for correlating extinct trackmakers and tracks, but two methods exist for making such correlations. The first involves trying to fill the gap by reconstructing the fleshy part of the foot, starting with either the skeleton or the track. Although many dinosaurs and other vertebrates have been 'fleshed out'²⁷, there have been no serious attempts to reconstruct feet either from the track up or from the foot skeleton down. The second approach involves the simple process of elimination, that is, matching tracks with trackmakers of the same size and geological age. □



Footprint morphology is not a simple reflection of foot skeletons with thin skin, as footprints of large dinosaurs such as *Hadrosaurus* (left) and *Tyrannosaurus* (right) show. Flesh and padding (grey areas) surround the bone. (Hallux restoration is omitted for clarity.)

were proposed in support of the idea that dinosaurs were warm-blooded, and athletic^{27,28}. However, there is little evidence for high-speed progression except among small- and medium-sized bipeds (mainly theropods)^{2,29}. Ichtnology has also forced debate about ceratopsian posture and locomotion, by showing that forelimbs were not placed as far from the midline as suggested by many reconstructions⁷. The use of footprints has also settled a debate about pterosaur locomotion by showing that they progressed quadrupedally on land^{8,30,31}.

A synthesis of all known brontosaur trackways (>400) reveals a pronounced shift from mixed gauge (narrow and wide) to predominantly wide gauge at the end of the Jurassic period¹⁵, presumably reflecting a major trend in sauropod evolution—that is, the rise of the brachiosaurids and titanosaurs. Another unexpected result is the recognition of considerable variation in heteropody (manus:pes foot area ratios vary from 1:5 to 1:2)¹⁵. Large manus sauropod trackways from the Upper Cretaceous have been identified as titanosaurid simply because they were the only family extant at that time. Such biostratigraphic processes of elimination, in combination with evi-

dence of gauge, heteropody and palaeobiogeography, help to differentiate the trackways of other sauropod families.

It is becoming clear that manus tracks of large Cretaceous ornithopods are variable, ranging from oval⁴ to bilobed³² and V-shaped³³. Such variation justifies the recognition of different ichnotaxa and increases the potential to differentiate between trackmakers at progressively lower taxonomic levels. In the absence of foot bones, sophisticated analyses of limb elements may be required to determine whether mammals were digitigrade or plantigrade³⁴, whereas tracks can distinguish these and unguligrade types unequivocally.

Megalosaurid tracks in Late Jurassic deposits from North America, Europe and Asia offer insight into a family that is poorly known from osteological remains¹⁶. The footprints are the largest theropod tracks known from the Jurassic period (up to 70 cm in length) and show that the megalosaurids moved with a short step and primitive gait (wide trackway and low pace angulation; Fig. 2)³⁵. Osteological evidence confirms that megalosaurs were primitive, and long-bodied, with relatively short, bandy legs³⁶.

Ichtnology is also useful in study of social behaviour^{3,4,37}. It is now known that large herbivorous dinosaurs (especially brontosaurs and ornithopods) often travelled in herds, and that both groups acquired gregarious tendencies early in their history. From study of trackway samples of up to 80 or more individuals, we know that such herds were sometimes segregated into distinct age or size classes^{4,37}. Claims that brontosaurs are represented in the fossil record by a huge preponderance of adults apply only to skeletal remains. Many large footprint samples, from continental palaeoenvironments, are dominated by the tracks of juveniles¹⁵. Small sauropod tracks from island archipelago palaeoenvironments, however, may indicate the presence of dwarfed populations³⁸.

As tracks are found *in situ*, they offer palaeoecological insight into trackmaker–habitat relationships. Recurrent associations between particular track assemblages and sedimentary facies allow recognition of distinct tetrapod ichnofacies³⁹. The most striking example is perhaps the recurrent association of synapsid and arthropod (arachnid/scorpion) tracks in sand dune facies throughout much of the Permian, Triassic and Early Jurassic in North and South America and Europe. In contrast, non-eolian facies exhibit entirely different, and more diverse, tetrapod track assemblages. Our understanding of the palaeoecology of such Jurassic units as the Wingate, Navajo, Carmel and Summerville (Fig. 1) is based almost exclusively on footprints. There are also examples of lacustrine and lagoonal ichnofacies dominated by swim-tracks attributed to amphibians, turtles⁴⁰ and pterosaurs⁴¹.

There is a statistical correlation between brontosaur tracks and low-latitude, carbonate-evaporite substrates¹⁵, which is relevant to debates about sauropod metabolism and biogeography. In contrast, large Mesozoic ornithopods are associated with plant-rich, coal-bearing facies indicative of humid, well-vegetated habitats³. Such ichnofacies distribution patterns are empirically derived, repeatable results that describe the composition of the track record. Similarly, in the Triassic of North America, for example, the body fossil record is dominated by aquatic vertebrates with sparse representation of terrestrial faunas. The track record, however, compensates for this bias with good representation of terrestrial tetrapods^{4,5}.



Figure 2 This trackway of a megalosaurid (*Megalosaurus*) from the Late Jurassic stage of Turkmenistan is unique among theropod trackways in its irregularity and width of gauge, indicative of a primitive style of locomotion³⁵.

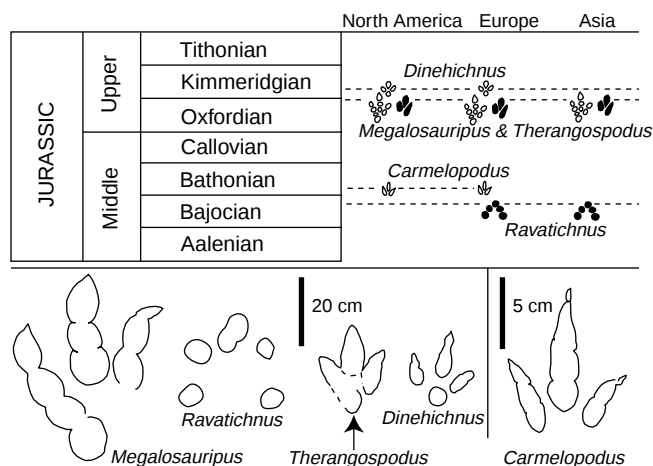


Figure 3 Ichonologists have recently reported correlations between five Middle and Upper Jurassic sites of North America, Europe and Asia, using dinosaur tracks. The sites are illustrated by their associated tracks at the right; the tracks are shown in more detail at the bottom. Few precise terrestrial correlations have been demonstrated, at this time, using dinosaur body fossils.

Calibrating the fossil record

Vertebrate tracks have definable distributions in space and time. Their potentially large geographic range, especially pertinent when considering birds and pterosaurs, makes them useful for biostratigraphy. For example, the large bird track *Magnoavipes* from the Lower Cenomanian of Texas⁴² is also known from deposits of the same age in Israel⁴³. The subdiscipline of ancient track stratigraphy (palichnostratigraphy), first introduced in the context of wide-ranging Pangaeian (Permo/Triassic/Early Jurassic) faunas⁴⁴, can now be extended to the Middle and Late Jurassic^{35,45}. Such correlations show widespread interchange of species between Europe and Eurasia after the break up of Pangaea. No such explicit (stage level) Middle or Upper Jurassic correlations have been made on the basis of dinosaur skeletal remains (Fig. 3 and Box 2).

Ichnology has also provided insight into the origin, evolution and extinction of several major groups and behaviours. Trackways have been used to suggest that the first tetrapods walked on land at about the time of the Silurian/Devonian transition, as much as 30–40 million years before the appearance of the oldest known skeletal remains^{46,47}, although this interpretation has been challenged⁴⁸. Because tracks are *in situ* they assume added importance in the debate about whether tetrapod 'walking' originated in fresh or salt water⁴⁹. Although not now regarded as tetrapod in origin, there is a Carboniferous acme zone of giant myriapod trails attributed to *Arthropleura*⁵⁰. Spiral burrows and digging traces also attest to the Permian origin of den-making activity of synapsids⁵¹.

In many areas we know more about Triassic archosaurian diversification from famous *Chirotherium* and *Chirotherium*-like tracks than we do from bones^{44,52,53}. Preliminary studies of Mesozoic bird tracks indicate a large-scale radiation of shorebirds in the early and mid Cretaceous as much as 40 million years before the appearance of skeletal remains of a postulated Charadriiform ancestor in the Late Cretaceous^{43,54}. Discovery of pterosaur track sites points to a Jurassic ichnological acme zone in association with marginal marine habitats and a preference of pterosaurs for terrestrial, lacustrine habitats in the Cretaceous⁴³ (Fig. 4).

There are also abundant tracks of hadrosaurs and ceratopsians, in what was previously defined as a 3-m gap below the Cretaceous/Tertiary (K/T) boundary. The gap is now reduced to a maximum of 37 cm (refs 3, 4), solving the debate about whether dinosaurs went extinct long before, or in conjunction with, the K/T transition.

The Tertiary and Quaternary track record, for example the diverse avian ichnofauna of the Swiss Oligocene/Miocene⁵⁵, also improves

Box 2 Globe-trotting dinosaurs

The distinctive Middle Jurassic track *Carmelopodus* from the Bathonian of Utah reveals the existence of a previously unknown species, with a short metatarsal IV, that also made tracks in Bathonian deposits in England⁴⁶. The enigmatic track *Ravaticchnus* has been correlated from Asia to Europe (Fig. 3)³⁵. Similarly, large 'megalo-saur' tracks (*Megalosauropus*) occur at the Oxfordian/Kimmeridgian (Upper Jurassic) transition in Utah, Arizona, New Mexico, Oklahoma, Portugal, Uzbekistan and Turkmenistan (Fig. 2)³⁵. Another distinctive dinosaur track type (*Therangospodus*) occurs in many of these samples³⁵ (Fig. 3). *Dinehichnus* is also correlated between North America and Europe⁶⁵. Such stage-level correlations exceed the resolution of only three land-vertebrate ages spanning 50 Myr⁶⁶. Track correlations tie footprint zones to stages with a duration of about 7 Myr each, and are as precise as any durations proposed for dinosaur skeletal taxa¹⁹.

our understanding of the palaeobiology of birds and mammals. The mammal-dominated, Pliocene australopithecine footprint site at Laetoli⁵⁶ pushed back the confirmed origins of hominid bipedalism from 3.0 to at least 3.6 million years⁵⁷, and provided evidence of a social group in which a juvenile was travelling with two adults⁵⁸. Late Pleistocene (Palaeolithic) hominid tracks from cave sites show a predominance of juveniles, without significant incidence of foot deformities, in association with bear, hyaena and (rare) fox tracks⁵⁹. In contrast, outdoor Mesolithic⁶⁰ and Neolithic⁶¹ sites contain more adult tracks, with higher incidence of foot abnormalities, in association with auroch, deer, unshod horse and bird footprints⁶¹. The use of the term 'trace fossil' for several categories of butcher cut marks on bone⁶² significantly broadens our concept of hominid ichnology.

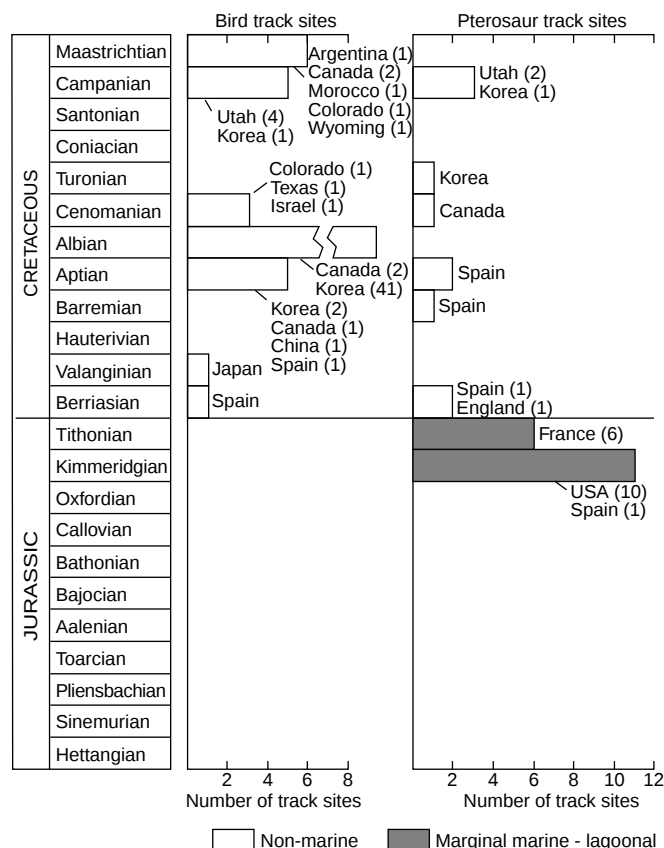


Figure 4 The stratigraphic distribution of bird and pterosaur tracks from 91 sites extends the temporal and spatial distribution of both groups⁴³.

Bias and completeness in the fossil record

Ichnology cannot solve the common taphonomic problem of missing foot skeletons entirely. The perception that vertebrate palaeontologists are waiting for ichnologists to find footprints to match their small sample of foot bones—what I call the ‘Cinderella’ syndrome—is largely unjustified. Often it is tetrapod ichnologists who are waiting for palaeontologists to find foot bones to fit into their large sample of tracks (an ‘overstocked shoe shop’ syndrome). This is because it is the feet that are missing from so many taxa as a result of taphonomic loss. Viewed from this perspective, the ichnological community has already found much of the ‘missing’ foot evidence and is establishing its own database. For example, the aforementioned Neolithic track site has yielded more than 150 hominid trackways, differentiated into males, females and children⁶¹. A comparable cache of skeletal remains would be considered quite exceptional.

So vertebrate ichnology is rapidly coming of age. Although vertebrate ichnology does not lend itself easily to the study of relationships through cladistic analysis, the field is highly consistent in comparing foot (manus and pes) morphology from taxon to taxon. This is not always true in the systematic study of bones, as the large number of invalid taxa, based on diverse undiagnostic elements, shows. Although there are many invalid track taxa also, there is a trend towards caution in ichnotaxonomy.

At a time when palaeontology has been aided by revolutions in such subdisciplines as molecular palaeontology and cladistics, the study of tetrapod tracks has also grown steadily and undergone its first true renaissance since its inception more than 150 years ago^{1–4,9,10}. This has largely been due to an increase in the discovery and documentation of new sites. This empirically derived database helps us define and measure spatial and temporal incompleteness in the skeletal record of terrestrial vertebrates. Thus, we can focus on which parts of the fossil record must be described ichnologically, osteologically or by both methods. This holistic view shows that many questions about foot morphology, posture, locomotion, behaviour, terrestrial vertebrate palaeoecology, stratigraphy and information biases cannot be fully understood without consideration of the data encoded in the track record. □

Martin G. Lockley is at the University of Colorado at Denver, Department of Geography, Geology, and Environmental Science, CB 172, PO Box 173364, Denver, Colorado 80217-3364, USA

- Gillette, D. D. & Lockley, M. G. (eds) *Dinosaur Tracks and Traces* (Cambridge Univ. Press, Cambridge, 1989).
- Thulborn, R. A. *Dinosaur Tracks* (Chapman & Hall, London, 1990).
- Lockley, M. G. *Tracking Dinosaurs* (Cambridge Univ. Press, Cambridge, 1991).
- Lockley, M. G. & Hunt, A. P. *Dinosaur Tracks and Other Fossil Footprints of the Western United States* (Columbia Univ. Press, New York, 1995).
- Lockley, M. G. & Hunt, A. P. in *Mesozoic Systems of the Rocky Mountain Region* (eds Caputo, M. V. et al.) 95–108 (SEPM, Society for Sedimentary Geology, Denver, 1994).
- Bakker, R. T. The dinosaur renaissance. *Sci. Am.* **232**, 58–72 and 77–88 (1975).
- Lockley, M. G. & Hunt, A. P. Ceratopsid tracks and associated ichnofauna from the Laramie Formation (Upper Cretaceous: Maastrichtian) of Colorado. *J. Vert. Paleontol.* **15**, 592–614 (1995).
- Lockley, M. G., Moratalla, J., Hunt, A. P., Schultz, R. J. & Robinson, J. W. The fossil trackway *Pteraiichnus* is pterosaurian, not crocodilian: implications for the global distribution of pterosaurs track. *Ichnos* **4**, 7–20 (1995).
- Buckland, W. *Bridgewater Treatise VI, Geology and Mineralogy Considered With Reference to Natural Theology* 3rd edn. Vol. I 552, Vol. II 143 (Pickering, London, 1858).
- Hitchcock, E. *The Ichnology of New England* (White, Boston, 1858).
- de Lapparent, A. F. Footprints of dinosaurs in the Lower Cretaceous of Vestspitsbergen-Svalbard. *Arbok Norsk. Polaristitut. for 1960*, 14–21 (1962).
- Currie, P. J., Nadon, G. & Lockley, M. G. Dinosaur footprints with skin impressions from the Cretaceous of Alberta and Colorado. *Can. J. Earth Sci.* **28**, 102–115 (1990).
- Olsen, P. E. & Galton, P. M. A review of the reptile and amphibian assemblages from the Stormberg of Southern Africa, with special emphasis on the footprints and the Age of the Stormberg. *Paleontol. Afr.* **25**, 87–110.
- Leonardi, G. *Annotated Atlas of South American Tetrapod footprints (Devonian to Holocene)*. Brazilia: República Federativa do Brasil, Ministério de Minas e Energia (Compania de Pesquisa de Roucourses Minerals, Brasília, 1994).
- Lockley, M. G., dos Santos, V. F., Meyer, C. A. & Hunt, A. P. (eds) Aspects of sauropod paleobiology. *Gaia: Revista de Geociências, Museu Nacional de Historia Natural* **10**, 1–279 (1994).
- Weishampel, D. B., Dodson, P. & Osmolska, H. *The Dinosauria* (Univ. California Press, Berkeley, Los Angeles, Oxford, 1990).
- Lucas, S. G. & Heckert, A. B. (eds) Early Permian footprints and facies. *New Mexico Mus. Nat. Hist. Sci. Bull.* **6**, 1–301 (1994).
- Dodson, P. & Dawson, S. D. Making the fossil record of dinosaurs. *Mod. Geol.* **16**, 13–15 (1991).

- Dodson, P. Counting dinosaurs: how many kinds were there? *Proc. Natl Acad. Sci. USA* **87**, 7608–7612 (1990).
- Farlow, J. O. & Brett-Surnam, M. K. (eds). *The Complete Dinosaur* 554–578 (Indiana Univ. Press, Bloomington, Indianapolis, 1997).
- Haubold, H. *Ichnia Amphibiorum et Reptiliorum fossilium. Handbuch Palaeoherpetologie* **8**, 124 (1971).
- Paul, G. The many myths, some old, some new of dinosaurology. *Mod. Geol.* **16**, 69–99 (1991).
- Allen, J. R. L. Subfossil mammalian tracks (Flandrian) in the Severn Estuary, S. W. Britain: mechanics of formation, preservation and distribution. *Phil. Trans. R. Soc. Lond. B* **352**, 481–518 (1997).
- Sarjeant, W. A. S. in *Dinosaur Systematics* (ed. Carpenter, K.) 299–307 (Cambridge Univ. Press, Cambridge, 1990).
- Nadon, G. C. & Issler, D. R. The compaction of floodplain sediments: timing, magnitude and implications. *Geosci. Can.* **24**, 37–43 (1997).
- Alexander, R., Mc, N. Estimates of speeds of dinosaurs. *Nature* **261**, 129–130 (1976).
- Czerkas, S. & Olsen, E. C. (eds) *Dinosaurs Past and Present* (Univ. Washington Press, Seattle, 1987).
- Russell, D. A. & Belland, P. Running dinosaurs. *Nature* **264**, 486 (1976).
- Farlow, J. O. Estimates of dinosaur speeds from a new trackway site in Texas. *Nature* **294**, 747–748 (1981).
- Padian, K. & Olsen, P. The fossil trackway *Pteraiichnus*: not pterosaurian but crocodilian. *J. Paleontol.* **58**, 178–184 (1984).
- Clark, J. M. et al. Foot posture in a primitive pterosaur. *Nature* **391**, 886–889 (1998).
- Moratalla, J. J. et al. A quadrupedal ornithopod trackway from the Early Cretaceous of La Rioja (Spain): inferences on gait and hand structure. *J. Vert. Paleontol.* **12**, 150–157 (1992).
- Wright, J. Ichnological evidence for the use of the forelimb in iguanodontid locomotion. *Palaeontology* (in the press).
- Carrano, M. T. Morphological indicators of foot posture in mammals: a statistical and biomechanical analysis. *Zool. J. Linn. Soc.* **121**, 77–104 (1997).
- Lockley, M. G., Meyer, C. A. & dos Santos, V. F. *Continental Jurassic* (ed. Morales, M.) 113–118, 137–140 (Museum of Northern Arizona, Flagstaff, 1996).
- Bakker, R. T. *Continental Jurassic* (ed. Morales, M.) 35–49 (Museum of Northern Arizona, Flagstaff, 1996).
- Lockley, M. G. Track records. *Nat. Hist.* **104**, 46–51 (1995).
- Dalla Vecchia, F. E. M. Jurassic and Cretaceous sauropod evidence in the Mesozoic carbonate ‘forms of the southern Alps and Dinards. *Gaia: Revista de Geociências, Museu Nacional de Historia Natural* **10**, 65–73 (1994).
- Lockley, M. G., Hunt, A. P. & Meyer, C. *The Paleobiology of Trace Fossils* (ed. Donovan, S.) 241–268 (Wiley, New York, 1994).
- Bernier, P. *Un Lagune Tropicale au Temps des Dinosaures* 1–136 (Centre National de la Recherche Scientifique, Museum de Lyons, 1985).
- Mazin, J.-M. et al. Le gisement de Crayssac (Tithonien inférieur, Quercy, Lot, France) découverte de piste de dinosaures en place et premier bilan ichnologique. *C.R. Acad. Sci.* **325**, 733–739 (1997).
- Lee, Y.-N. Bird and dinosaur footprints in the Woodbine Formation (Cenomanian), Texas. *Cretac. Res.* **18**, 859–864 (1997).
- Lockley, M. G. & Rainforth, E. C. in *Mesozoic Birds* (eds Chiappe, L. & Whitmer, L. M.) (Univ. Calif. Press, Berkeley, Los Angeles, Oxford, in the press).
- Haubold, H. & Katzung, G. Paleoeology and paleoenvironments of tetrapod footprints from the Rotliegend (Lower Permian) of central Europe. *Paleogeogr. Paleoclimatol. Paleocool.* **23**, 307–323 (1978).
- Lockley, M. G., Hunt, A. P., Paquette, M., Bilbey, S.-A. & Hamblin, A. Dinosaur tracks from the Carmel Formation, northeastern Utah: implications for Middle Jurassic paleoecology. *Ichnos* **5**, 255–267 (1998).
- Warren, A., Jupp, R. & Bolton, B. Earliest tetrapod trackway. *Alcheringa* **10**, 183–186 (1986).
- Stosell, I. The discovery of a new Devonian tetrapod trackway in SW Ireland. *J. Geol. Soc. Lond.* **152**, 407–413 (1995).
- Clack, J. A. Devonian tetrapod trackways and trackmakers; a review of the fossils and footprints. *Paleogeogr. Paleoclimatol. Paleocool.* **130**, 227–250 (1997).
- Coates, M. I. & Clack, J. A. Romer’s gap: tetrapod origins and terrestriality. *Bull. Mus. Nat. Hist. Nat. Paris* **17**, 373–388 (1995).
- Briggs, D. E. G., Plint, A. G. & Pickerill, R. K. *Arthropoda* trails from the Westphalian of eastern Canada. *Paleontol.* **27**, 843–855 (1984).
- Smith, R. M. H. Helical burrow casts of therapsid origin from the Beaufort Group (Permian) of South Africa. *Paleogeogr. Paleoclimatol. Paleocool.* **60**, 155–170 (1987).
- Demathieu, G. & Haubold, H. Evolution und Lebensgemeinschaft terrestrischer Tetrapoden nach ihren Fährten in der Trias. *Freiberg. Forschung. C* **298**, 51–72 (1974).
- Peabody, F. Reptile and amphibian trackways from the Lower Triassic Moenkopi Formation of Arizona and Utah. *Bull. Dep. Geol. Sci. Univ. Calif.* **27**, 295–468 (1948).
- Lockley, M. G., Yang, S.-Y., Matsukawa, M., Fleming, F. & Lim, S.-K. The track record of mesozoic birds: evidence and implications. *Phil. Trans. R. Soc. Lond. B* **336**, 113–134 (1992).
- Weidmann, M. & Reichel, M. Trace de pattes d’oiseaux dans la Molasse Suisse. *Eclogae Geol. Helv.* **72**, 953–971 (1992).
- Leakey, M. D. & Hay, R. L. Pliocene footprints in the Laetoli beds at Laetoli, northern Tanzania. *Nature* **278**, 317–323 (1979).
- Day, M. H. & Williams, E. H. Laetoli Pliocene hominid footprints and bipedalism. *Nature* **286**, 385–387 (1980).
- Hay, R. L. & Leakey, M. D. The fossil footprints of Laetoli. *Sci. Am.* **246**, 50–55 (1982).
- Pales, L. Les empreintes de Pieds Humains dans les Cavernes. *Arch. Inst. Paleontol. Hum.* **36**, 1–166 (1976).
- Aldhouse-Green, S. H. R. et al. Prehistoric human footprints from the Severn Estuary at Uskmouth and Magor Pill, Gwent, Wales. *Archaeol. Cambrensis* **141**, 14–55 (1995).
- Roberts, G., Gonzalez, S. & Huddart, D. Intertidal Holocene footprints and their archaeological significance. *Antiquity* **70**, 647–651 (1996).
- Noe-Nygaard, N. Ecological, sedimentary and geochemical evolution of the late Glacial to postglacial Amose lacustrine basin, Denmark. *Fossils Strata* **37**, 1–436 (1995).
- Lockley, M. G. & Hunt, A. P. A track of the giant theropod dinosaur *Tyrannosaurus* from close to the Cretaceous/Tertiary boundary, Northern New Mexico. *Ichnos* **3**, 213–218 (1994).
- Guerin, C. & Demathieu, G. Empreintes et piste de Rhinocerotidae (mammalia, Perissodactyla) du Gisement Pliocene terminal de Laetoli (Tanzanie). *Geobios* **26**, 497–513 (1993).
- Lockley, M. G., dos Santos, V. F., Meyer, C. A. & Hunt, A. P. A new dinosaur tracksite in the Morrison Formation, southeastern Utah. *Mod. Geol.* **23**, 317–330 (1998).
- Lucas, S. G. Vertebrate biochronology of the Jurassic-Cretaceous boundary, North American western interior. *Mod. Geol.* **18**, 371–390 (1993).

Acknowledgements. I thank D. Norman and J. Clack for reading this manuscript and J. Snider for help with graphics.

Enhanced long-term potentiation and impaired learning in mice with mutant postsynaptic density-95 protein

Martine Migaud^{*†}, Paul Charlesworth^{*†}, Maureen Dempster^{*†}, Lorna C. Webster^{*†}, Ayako M. Watabe[‡], Michael Makhinson[‡], Yong He[§], Mark F. Ramsay[†], Richard G. M. Morris[†], John H. Morrison[§], Thomas J. O'Dell[‡] & Seth G. N. Grant^{*†}

^{*} Centre for Genome Research, and [†] Centre for Neuroscience, University of Edinburgh, Roger Land Building, West Mains Road, Edinburgh EH9 3JQ, UK

[‡] Department of Physiology and The Brain Research Institute, School of Medicine, University of California at Los Angeles, Los Angeles, California 90095, USA

[§] Fishberg Research Centre for Neurobiology, Mount Sinai School of Medicine, Box 1065, One Gustave Levy Place, New York, New York 10029-6574, USA

Specific patterns of neuronal firing induce changes in synaptic strength that may contribute to learning and memory. If the postsynaptic NMDA (*N*-methyl-D-aspartate) receptors are blocked, long-term potentiation (LTP) and long-term depression (LTD) of synaptic transmission and the learning of spatial information are prevented. The NMDA receptor can bind a protein known as postsynaptic density-95 (PSD-95), which may regulate the localization of and/or signalling by the receptor. In mutant mice lacking PSD-95, the frequency function of NMDA-dependent LTP and LTD is shifted to produce strikingly enhanced LTP at different frequencies of synaptic stimulation. In keeping with neural-network models that incorporate bidirectional learning rules, this frequency shift is accompanied by severely impaired spatial learning. Synaptic NMDA-receptor currents, subunit expression, localization and synaptic morphology are all unaffected in the mutant mice. PSD-95 thus appears to be important in coupling the NMDA receptor to pathways that control bidirectional synaptic plasticity and learning.

In the mammalian hippocampus and other brain structures involved with memory formation, certain patterns of synaptic activity result in long-lasting modifications of the efficiency of synaptic transmission. In the CA1 region of the hippocampus, a train of low-frequency stimulation produces a reduction in synaptic efficacy (LTD) and high-frequency stimulation produces an increase (LTP)^{1,2}. Mechanisms proposed to underlie these modifications of synaptic strength include changes in the sensitivity of postsynaptic transmitter receptors, activation of previously silent receptors, generation of retrograde messengers to the presynaptic terminal, structural changes in dendritic spines, and activation of transcription in the nucleus^{1,2}. Although activation of the NMDA receptor initiates these complex events, its interactions with postsynaptic proteins and the signalling pathways immediately downstream from it are poorly understood.

The NMDA receptor is formed by the assembly of a common NR1 subunit with one or more of four different NR2 subunits, NR2A–D (refs 3, 4). The NR2 subunits have long cytoplasmic carboxy-terminal domains which contain sites for phosphorylation⁵ and interaction with cytoplasmic proteins. *In vitro*, the C terminus of NR2 subunits binds to PSD-95/SAP90, chapsyn-110/PSD-93, and other related members of the membrane-associated guanylate kinase (MAGUK) family^{6–9}. PSD-95/SAP90 (refs 10, 11) is an abundant postsynaptic density protein and contains several domains that participate in protein–protein interactions, including three PDZ/DHR (for PSD-95, Dlg, ZO-1/Dlg-homologous region) domains, an SH3 (Src-homology-3) domain and a guanylate kinase (GK)-homology domain. The second PDZ domain (PDZ2) can bind to the C terminus of the NR2 subunits^{6,9}. PSD-95 has no detectable enzymatic activity and acts as an adapter molecule through protein–protein interactions mediated by the discrete domains. Although the function of PSD-95 at the synapse is unknown, in fibroblasts it mediates the localization of the NMDA receptor to focal clusters, indicating that PSD-95 may be required for the localization of such receptors to synapses⁷.

To determine the function of PSD-95 in the brain, we character-

ized mice carrying a targeted mutation in the PSD-95 gene. Our results provide evidence that PSD-95 is important in signal transduction. NMDA-receptor-mediated synaptic plasticity was dramatically altered, with synapses becoming inappropriately strengthened after stimulation by a wide range of frequencies. The learning of PSD-95-mutant mice was impaired, supporting predictions of neural-network models that depend on bidirectional synaptic plasticity to mediate learning and memory.

PSD-95-mutant mice

The first two PDZ domains of PSD-95 bind to the NMDA-receptor subunits 2A (NR2A) and 2B (NR2B)^{6,9}, to K⁺ channels¹² and to neuronal nitric oxide synthase¹³ *in vitro*. We generated mice carrying a targeted mutation in the PSD-95 gene that leaves the first two PDZ domains intact by introducing a stop codon into the PDZ3 domain and replacing downstream sequences with an internal ribosome entry site that drives a β -galactosidase reporter gene (Fig. 1a). The targeting construct was electroporated into embryonic stem (ES) cells and Southern blot analysis indicated that homologous recombination had occurred in 5 clones out of 75 analysed (6.7%) (Fig. 1b). Three ES cell lines were microinjected into C57BL/6 blastocysts and two clones generated germline chimaeras, which were crossed onto the MF1 genetic background. No biochemical, anatomical, electrophysiological or behavioural differences were found between mice generated from two independent targeted clones. Heterozygous mice were intercrossed and the genetic status of the mice was determined by genomic Southern blotting and polymerase chain reaction (PCR) (Fig. 1b). From 70 litters, 602 mice were genotyped at weaning: 191 (32%) were wild type, 316 (52%) were heterozygotes, and 95 (16%) were homozygotes, demonstrating a distortion of the expected mendelian ratio ($\chi^2 = 32.1$, $P < 0.0001$). Runting was evident in 42% of viable homozygotes, which then recovered and reached the size of their wild-type littermates at ~6 weeks of age. Runting seemed to result from inadequate competition for feeding with wild-type siblings, because homozygous pups fostered to wild-type mothers were of normal size. Homozygous males and females were fertile and showed

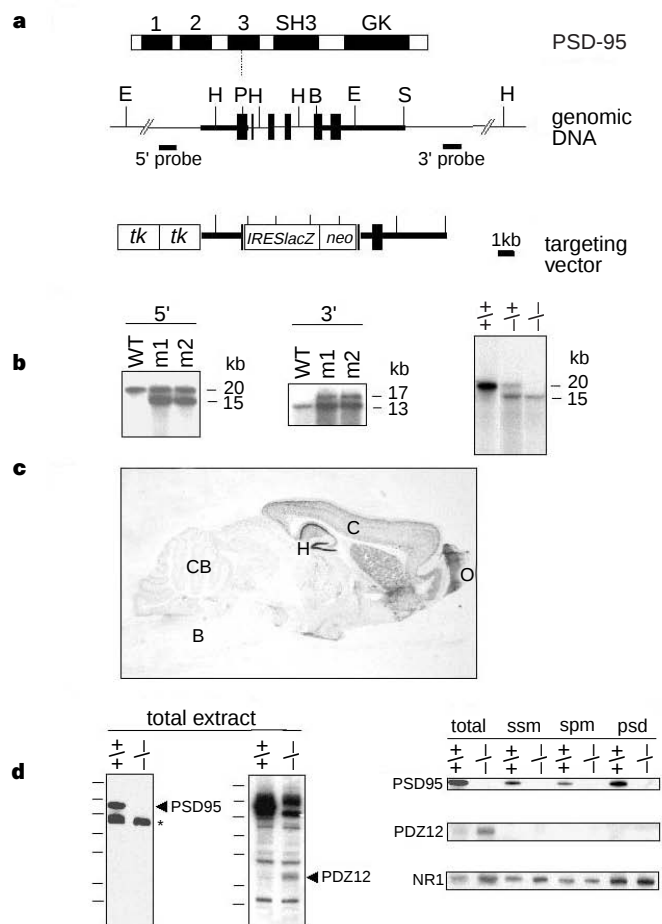


Figure 1 Targeted disruption of the PSD-95 gene. **a**, Top: PSD-95 protein with three PDZ domains (labelled 1, 2, 3), an Src homology 3 (SH3) domain and a guanylate kinase homology region (GK). Middle: PSD-95 gene with restriction-enzyme cutting sites (E, *EcoRI*; H, *HindIII*; P, *PvuII*; B, *BamHI*; S, *SacI*); black boxes, exons, thick horizontal lines, homology regions used in the targeting vector; Southern blot probes are indicated. Bottom: targeting vector. *tk*, thymidine kinase gene; *IRES*, internal ribosome-entry sequence; *lacZ*, β -galactosidase gene; *neo*, neomycin-resistance gene. The dotted vertical line from the PDZ3 domain (upper) to the *PvuII* site (middle) represents the position of the stop codon inserted into PDZ3. **b**, Left and centre, genomic Southern blot from ES cell clones (m1 and m2) digested with *EcoRI* and *HindIII*, hybridized with 5' and 3' probes, respectively. Right: genomic Southern blot of *EcoRI*-digested DNA from littermates of a heterozygote intercross and probed with the 5' probe. Wild type (WT) (+/+), heterozygote (-/+) and homozygote (-/-) are indicated. **c**, Expression pattern of PSD-95 using X-gal-stained sagittal brain section. B, brainstem; H, hippocampus; C, cortex; CB, cerebellum; O, olfactory bulb. **d**, Immunoblots of PSD-95 protein in total forebrain extracts (left, middle panels) and in synaptosome subfractions (right panels) of wild-type mice (+/+) and PSD-95 mutants (-/-). Left: full-length PSD-95 in wild-type is indicated; asterisk, nonspecific band; size markers correspond to M, 205K, 112K, 87K, 69K, 56K, 39K, 34K. Middle: the same blot was stripped and reprobed with an antibody recognizing PSD-95 N-terminal to the PDZ3 domain. The band labelled PDZ12 is found only in extracts of PSD-95 mutant mice and corresponds to the N-terminal domain expressed in the mutants. Right: extracts from synaptosome (ssm), synaptic plasma membrane (spm) and postsynaptic density (psd) fractions from wild-type and mutant mice immunoblotted with antibodies recognizing full-length PSD-95 (upper strip), N-terminal PDZ12 (middle strip) and NR1 (lower strip). The PSD-95^{PDZ12} (PDZ12) protein was not detected in spm or psd fractions even at very long exposures.

no sign of seizure, tremor or ataxia or of neurological abnormality. The expression of PSD-95 was shown by X-gal staining of brains of mutant mice (Fig. 1c) to be high in the forebrain, particularly in CA1 pyramidal neurons of the hippocampus and in granule cells in the dentate gyrus, in agreement with studies of messenger RNA *in situ*^{6,13}.

To determine the effects of mutating the PSD-95 protein we immunoblotted whole extracts from forebrain, synaptosomes, and synaptosome subfractions with antibodies raised against PSD-95 (Fig. 1d). Using three separate antibodies against PSD-95, a 95K band (corresponding to a relative molecular mass of 95,000) was detected for wild-type mice which was absent in homozygote mice. Immunoblotting with antibodies against the amino terminus of PSD-95 detected a 40K band in whole-brain extracts from homozygote mice, which corresponds to PDZ domains 1 and 2 and is now referred to as PSD-95^{PDZ12}. The PSD-95^{PDZ12} band was not detected in synaptosomes, in synaptic plasma membranes or in postsynaptic density fractions of homozygote mice, in contrast to full-length PSD-95 which was readily detectable in these fractions in wild-type mice (Fig. 1d). Control immunoblotting with antibodies specific for the NMDA R1 subunit (NR1) indicated that the distribution of NR1 in these fractions was the same as in wild-type mice. These results show that PSD-95^{PDZ12} does not localize to postsynaptic densities or to synaptic plasma membranes, but without a suitable antibody we cannot confirm this by immunohistochemistry. In PSD-95-mutant mice, no binding of PSD-95^{PDZ12} to the NMDA receptor could be detected by immunoprecipitation assay, although PSD-95 bound to the NMDA receptor from wild-type mice (data not shown). The amount of PSD-95^{PDZ12} found in PSD-95 mutants was ~10-fold less than full-length PSD-95 in wild-type, and as RNase protection assays indicated that the level of mRNA in each was comparable (data not shown), the reason why we detected no interaction between PSD-95^{PDZ12} and the NMDA receptor is probably because PSD-95^{PDZ12} is not localized to synapses. This is consistent with data showing that the C-terminal domains of Dlg are required for membrane and subcellular localization¹⁴.

Several other related neuronal MAGUK proteins have been described in addition to PSD-95, including chapsyn-110/PSD-93, SAP102 and SAP97 (refs 7, 8, 13, 15). The levels of these proteins were unaltered in our mutant mice (data not shown). We showed that chapsyn-110/PSD-93 could still interact with the NMDA receptor in PSD-95-mutant mice by immunoprecipitation and by immunoblotting the complex with antibodies against the NMDA receptor (data not shown). We conclude that the NMDA receptor in our PSD-95-mutant mice is bound by another MAGUK protein.

Synaptic localization of NMDA receptors

We next investigated the expression and localization of the NMDA receptor. The total amount of NR1, NR2A and NR2B proteins was the same in wild-type and mutant mice, and immunoprecipitation of each subunit and immunoblotting with the other subunits revealed no change in the composition of the receptor complexes (data not shown). These results indicate that PSD-95 is unlikely to play a role in regulating the stoichiometry of the NMDA-receptor subunit interaction *in vivo*. PSD-95 is highly expressed in the CA1 region of the hippocampus (Fig. 1c). Using light and electron microscopy, we found that the intensity of the Nissl staining pattern (data not shown) and the distribution of the dendritic marker MAP2 and of the synaptic-terminal marker synaptophysin were the same in the homozygous mutant and wild-type mice, indicating that the mutation does not affect cell density or cytoarchitectonic patterns (Fig. 2a–d). NR1 was widely distributed in soma and dendrites of CA1 in both mutant and wild-type mice and appeared similar in the light microscope (Fig. 2e, f). Electron microscopy revealed no changes in the morphology of these asymmetric synapses nor any difference in synapse density (Fig. 2i, j). Immunogold studies of NR1 in CA1 stratum radiatum detected no difference between the mutant and wild-type mice in the character-

istics or density of NR1 immunogold labelling of asymmetric synapses (Fig. 2g, h). In both sets of mice, NR1 was present primarily in the synaptic cleft and within the postsynaptic specialization, with minimal labelling in the neighbouring cytoplasm and no significant presynaptic labelling. Although these techniques are not quantitative, we found no major difference between mutant and wild-type mice in hippocampus CA1 synapse architecture or in NMDA-receptor localization. The localization of the NMDA receptor to the synapses of PSD-95 mutants by electron microscopy was confirmed by electrophysiological assay (see below).

NMDA-receptor function in mutant mice

We investigated the NMDA currents in primary cultured neurons derived from PSD-95-mutant and wild-type mice. Peak currents elicited by brief application of 250 μ M NMDA/25 μ M glycine in magnesium-free medium were not significantly different ($3,907 \pm 285$ pA, $n = 36$ for PSD-95 mutant; $3,204 \pm 328$ pA, $n = 22$ for wild type; $t(56) = 1.578$, $P = 0.12$). Furthermore, the NMDA-receptor current–voltage relation did not change in PSD-

95-mutant neurons in the presence and absence of magnesium (Fig. 3a, b). We investigated the synaptic expression of NMDA receptors in hippocampal slices using whole-cell voltage-clamp recordings to examine the NMDA-receptor-mediated component of excitatory postsynaptic currents (EPSCs) in CA1 pyramidal cells: we found no difference in the results from cells of wild-type and mutant animals (Fig. 3c). At a holding potential of -80 mV, where the NMDA-receptor ion channels are blocked by Mg^{2+} , EPSCs in CA1 pyramidal cells from wild-type animals ($n = 15$ cells from 4 animals) and PSD-95 mutants ($n = 18$ cells from 5 animals) have a similar ($t(7) = 0.49$; not significant), small NMDA-receptor-mediated component. The NMDA-receptor-mediated component of EPSCs from wild-type and PSD-95 mutant cells was also the same ($t(7) = 0.27$; not significant) at a holding potential of $+40$ mV, where the NMDA-receptor-mediated component of the EPSCs is larger owing to the relief of the Mg^{2+} block of the channel. These results show that NMDA-receptor channel properties are unaffected and that the receptors are normally localized to synapses in PSD-95-mutant mice. Moreover, field potential recordings showed that the maximal α -amino-3-hydroxy-5-methyl-4-isoxazole propionate (AMPA)-receptor-mediated field excitatory postsynaptic potential (fEPSP) amplitude evoked in wild-type and mutant slices was not significantly different (wild type: 8.6 ± 0.5 mV, $n = 69$ slices, 13 animals; PSD-95 mutant: 7.8 ± 0.4 mV, $n = 70$ slices, 15 animals), indicating that there is no gross disruption of synaptic transmission in our PSD-95-mutant mice.

We also compared paired-pulse facilitation, a short presynaptic form of synaptic plasticity, in the two sets of mice (Fig. 3d). We found that paired-pulse facilitation at inter-pulse intervals of 20, 50, 100 and 200 ms was significantly greater in slices from PSD-95-mutant mice ($P < 0.05$). As PSD-95 is thought to be a postsynaptic protein in forebrain synapses¹⁶, this enhanced paired-pulse facilitation may be due to postsynaptic changes¹⁷, or to retrograde signalling to the presynaptic terminal through PSD-95 interaction with neuroligin and neurexin¹⁸ (as shown in Fig. 6a), or to some other effect such as an alteration in presynaptic K^+ -channel function¹⁹.

Synaptic plasticity in PSD-95 mutants

At many synapses in the brain, activation of postsynaptic NMDA receptors triggers complex multicomponent signalling pathways that can produce persistent changes in synaptic strength, such as LTP and LTD^{1,2}. We therefore investigated whether PSD-95 might have a role in NMDA-receptor signalling by examining NMDA-receptor-dependent forms of synaptic plasticity in hippocampal slices from PSD-95-mutant mice. A conventional high-frequency stimulation protocol (100 Hz) that induces NMDA-receptor-dependent LTP produced a significantly larger potentiation of synaptic transmission in slices from PSD-95-mutant mice ($t(12) = 2.59$, $P < 0.025$; Fig. 4a). Immediately after high-frequency stimulation, the potentiation of synaptic transmission was similar in slices from wild-type and PSD-95-mutant mice; however, although synaptic strength remained at a steady-state potentiated level in wild-type slices (fEPSPs were potentiated to $193.6 \pm 9.3\%$ of baseline 60 min after high-frequency stimulation; $n = 7$), synaptic strength continued to grow over the next 30 min in slices from PSD-95-mutant mice (fEPSPs potentiated to $242.8 \pm 17.5\%$ of baseline; $n = 7$). To find the full extent of the enhanced LTP and the maximal LTP that could be generated in slices from PSD-95-mutant mice, we induced saturating levels of LTP with multiple trains of 100-Hz stimulation. As shown in Fig. 4b, LTP saturated at a much higher potentiated level in slices from mutant mice ($298.4 \pm 39.0\%$ of baseline; $n = 5$) compared with wild-type slices ($174.8 \pm 14.0\%$ of baseline; $n = 4$, $t(7) = 2.98$, $P < 0.05$).

We next quantified LTP induced by trains of low-frequency presynaptic stimulation. A 900-pulse train of 5-Hz stimulation, a near-threshold protocol for LTP induction²⁰, induced a small amount of LTP in slices from wild-type mice; fEPSPs were potentiated to

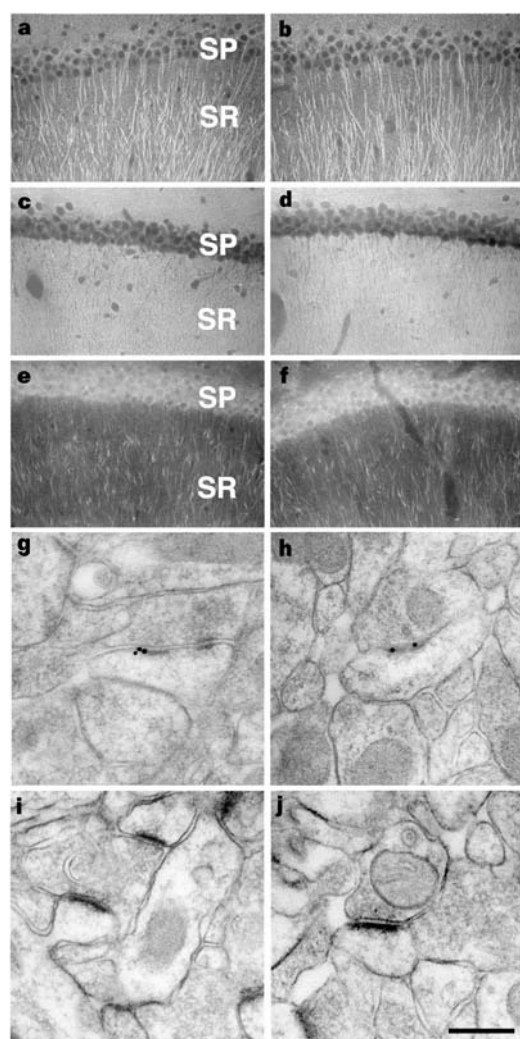


Figure 2 Anatomy and NR1 localization in hippocampus CA1 region of mice. **a, c, e, g, i**, Wild-type mice; **b, d, f, h, j**, PSD-95 mutant mice. **a, b**, MAP2 immunofluorescence; **c, d**, synaptophysin immunofluorescence; **e, f**, NR1 immunofluorescence; all photomicrographs are from a similar field of CA1, with the pyramidal cell layer (SP) at the top of the field, and the remaining two-thirds of the field occupied by the stratum radiatum (SR). **g, h**, Immunogold localization of NR1. Note that NR1 is localized at asymmetrical synapses in mutant and wild type mice. **i, j**, Standard morphology of the ultrastructure, with axospinous asymmetrical synapses shown in each case. Scale bar in **j** represents: for **a–f**, 70 μ m; for **g–j**, 0.25 μ m.

126.5 ± 7.3% of baseline after 45 min ($n = 7$) (Fig. 4c). In slices from PSD-95-mutant mice, synaptic transmission was potentiated to near saturation (fEPSPs were 278.4 ± 17.4% of baseline; $n = 6$) (Fig. 4c). The NMDA-receptor antagonist 2-amino-5-phosphonopivalic acid (AP5) blocked 5-Hz-stimulation-induced LTP in both wild-type and

PSD-95-mutant mice (fEPSPs in slices from PSD-95-mutant mice were 110.2 ± 11.7% of baseline ($n = 5$) in 100 μM DL-AP5 and 98.5 ± 3.4% of baseline in slices from wild-type mice ($n = 5$)), indicating that the enhanced LTP in slices from PSD-95-mutant mice is not due to upregulation of NMDA-receptor-independent

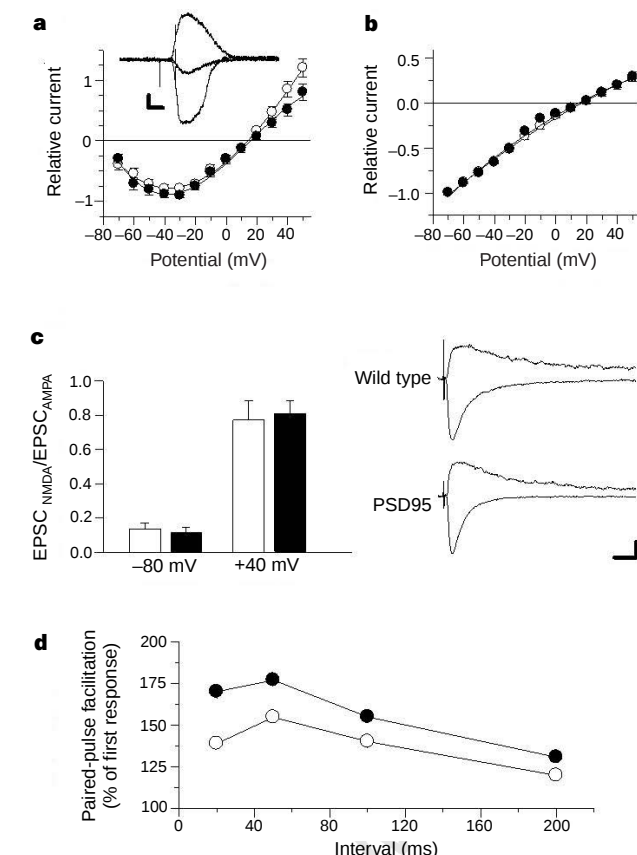


Figure 3 NMDA-receptor currents and synaptic physiology. **a**, Current-voltage relationship for NMDA (250 μM)-receptor-gated currents recorded in the presence of magnesium (1 mM) (white circles, wild type ($n = 10$); filled circles, PSD-95 mutant ($n = 16$)). Insert, currents recorded from a single PSD-95 mutant cell at -70 mV (middle trace), -20 mV (lower trace), and +50 mV (upper trace). Calibration bars are 50 ms and 100 pA. **b**, Current-voltage relationship for NMDA (250 μM)-receptor-gated currents recorded in magnesium-free solution (white circles, wild type ($n = 12$); filled circles, PSD-95 mutant ($n = 24$)). **c**, Synaptic NMDA-receptor currents. The bars show the magnitude of the NMDA-receptor-mediated component EPSCs (normalized to the magnitude of the AMPA-receptor-mediated component; see Methods). At postsynaptic holding potentials of both -80 mV and +40 mV, the NMDA-receptor-mediated component of the EPSCs in CA1 pyramidal cells from PSD-95 mutants (black bars) is indistinguishable from that seen in pyramidal cells from wild-type animals (white bars). Traces show representative EPSCs elicited in a wild-type cell (top) and a PSD-95-mutant cell (bottom) at +40 mV (outward-going currents) and -80 mV (inward-going currents). Each trace is the average of four EPSCs elicited at each holding potential. Calibration bars indicate 20 ms and 50 pA. **d**, Paired-pulse facilitation of fEPSPs was measured using pairs of presynaptic fibre stimulation pulses separated by 20, 50, 100 and 200 ms. Note the larger facilitation ($P < 0.05$) seen in slices from PSD-95-mutant animals (filled symbols) compared to wild type (open symbols). Each point is the mean ± s.e.m. (error bars are smaller than point used to plot the mean value).

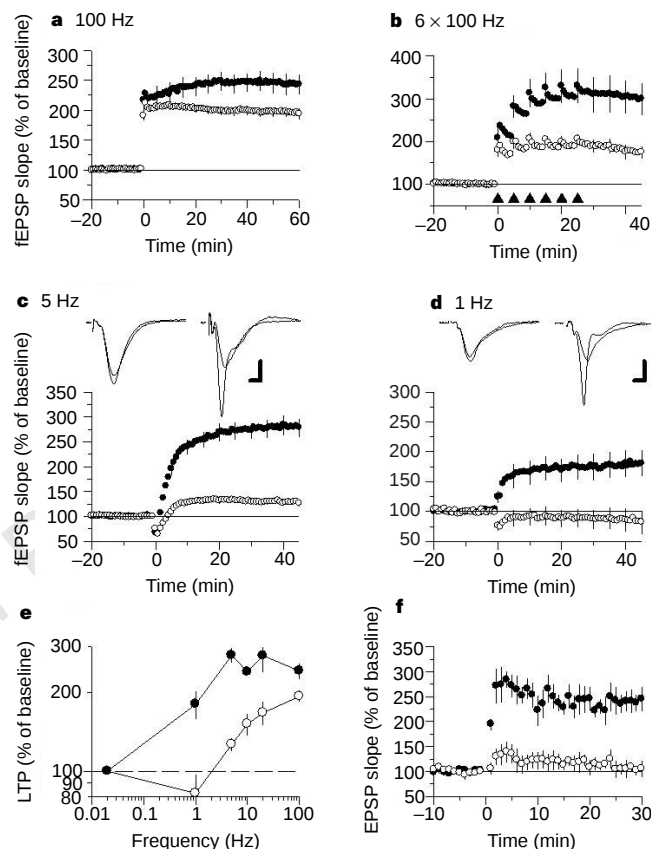


Figure 4 Frequency-dependent changes in synaptic strength. **a**, High-frequency stimulation-induced LTP was elicited using two trains of 100-Hz stimulation delivered at time zero. Note the larger potentiation seen 60 min after 100-Hz stimulation in slices from PSD-95 mutant animals (black symbols) compared with slices from wild-type animals (white symbols). **b**, Single 1-s trains of 100-Hz stimulation were delivered six times (at points indicated by black triangles) starting at time zero. Synaptic transmission in wild-type slices (white symbols) shows no further increase after the second train; transmission in slices for PSD-95 mutants continues to increase and saturates at a much higher level. **c**, **d**, 900 stimulation pulses were delivered at time zero at either 5 Hz (**c**) or 1 Hz (**d**). Although these stimulation protocols produce small amounts of either LTP (5 Hz) or LTD (1 Hz) in slices from wild-type animals (white symbols), both stimulation protocols induce large LTP in slices from PSD-95 mutant animals (black symbols). The fEPSPs in **c** were recorded during baseline and 45 min after 5-Hz stimulation in wild-type (left) and PSD-95 mutant animals (right). The fEPSPs in **d** were recorded before and 45 min after 1-Hz stimulation in slices from wild-type (left, smaller response is after 1-Hz stimulation) and PSD-95 mutant animals (right, larger response is after 1-Hz stimulation). Calibration bars in **c** and **d** indicate 2 mV and 2 ms. **e**, Summary of the ability of different frequencies of synaptic stimulation to induce persistent changes in synaptic strength in hippocampal slices from wild-type (white symbols) or PSD-95 mutant (black symbols) animals; 0.02 Hz, which corresponds to the stimulation frequency used to monitor synaptic transmission throughout the experiments, had no effect on synaptic transmission in slices from wild-type and PSD-95 mutant animals. The line at 100% of baseline corresponds to no lasting change in synaptic strength. **f**, EPSPs evoked once every 30 s were recorded intracellularly from individual CA1 pyramidal cells. Following 10 min of baseline recording, 30 EPSPs (evoked at 0.5 Hz) were each paired with a 150-ms, 1.0-nA depolarizing current pulse delivered through the recording electrode starting 10 ms after the start of an EPSP. Data are from 12 wild-type cells ($n = 4$ animals; white symbols) and 11 PSD-95 mutant cells ($n = 4$ animals; black symbols).

forms of LTP. A large enhancement of LTP in PSD-95-mutant slices was also stimulated by intermediate frequencies. In wild-type slices, 900 pulses of 10 and 20 Hz induced potentiations of $151.6 \pm 14.8\%$ ($n = 4$) and $167.4 \pm 17.7\%$ of baseline ($n = 6$), respectively, whereas fEPSPs from PSD-95-mutant mice were potentiated to $240.2 \pm 4.6\%$ ($n = 5$) and $277.3 \pm 37.5\%$ ($n = 6$) of baseline, respectively. Finally, we used a low-frequency protocol (1 Hz, 900 pulses) that induces LTD in hippocampal slices²¹ from young wild-type mice and induces a small and nonsignificant depression of synaptic transmission in slices from adult wild-type mice (fEPSPs were $82.2 \pm 19.6\%$ of baseline ($n = 4$, $t(3) = 0.91$), compared with pre-1 Hz baseline) (Fig. 4d). In response to this stimulation protocol, slices from adult PSD-95-mutant mice showed a significant LTP (fEPSPs were $180.4 \pm 21.8\%$ of baseline; $n = 5$, $t(4) = 3.69$, $P < 0.025$) (Fig. 4d). As summarized in Fig. 4e, these results indicate that the frequency sensitivity of LTP induction in the CA1 region of the hippocampus is altered in PSD-95-mutant mice.

To determine whether the enhanced LTP was due to altered inhibitory synaptic transmission, we tested potentiation in the presence of GABA inhibitors. Although 50 pulses of stimulation at 2.5 Hz elicited little LTP in slices from wild-type mice (fEPSPs were potentiated to $126.5 \pm 8.3\%$ of baseline; $n = 7$), we found that synaptic transmission was potentiated to $232.76 \pm 30.4\%$ of baseline in slices from PSD-95-mutant mice ($n = 3$, $t(8) = 4.93$, $P < 0.005$, compared to wild-type). When inhibitory synaptic transmission was blocked by 50–100 μM picrotoxin, 50 pulses of 2.5 Hz still elicited a significantly ($t(4) = 2.99$, $P < 0.025$) larger LTP in slices from PSD-95-mutant mice (fEPSPs were potentiated to $271.9 \pm 31.1\%$ ($n = 3$) of baseline in slices from PSD-95 mutants, and $175.6 \pm 8.3\%$ ($n = 3$) of baseline in slices from wild-type mice). In addition, when inhibitory synaptic transmission was blocked by 100 μM picrotoxin and intracellular microelectrodes were used to record EPSPs, we found that pairing low-frequency presynaptic fibre stimulation (0.5 Hz) with postsynaptic depolarization induced significantly larger LTP in cells from PSD-

95 mutants (30 min post-pairing EPSPs were $105 \pm 11.8\%$ of baseline in wild-type cells and $245 \pm 25.1\%$ of baseline in PSD-95 mutant cells; $t(6) = 4.64$, $P < 0.005$; Fig. 4f). Therefore, the enhanced induction of LTP in PSD-95 mutants does not seem to arise from alterations in inhibitory synaptic transmission. Moreover, the larger potentiation induced by pairing low-frequency presynaptic fibre stimulation with postsynaptic depolarization in PSD-95 mutants indicates that changes in paired-pulse facilitation or postsynaptic excitability are unlikely to underlie the LTP enhancement seen in PSD-95 mutants.

Learning and memory in PSD-95 mutants

Neural-network models of learning and memory incorporating bidirectional synaptic plasticity (LTP and LTD) predict that major alterations in the frequency function should be deleterious to learning. We therefore tested spatial learning in a watermaze, which is dependent on hippocampus NMDA-receptor function^{22,23}. Adult littermate wild-type ($n = 9$) and homozygous ($n = 12$) mice swam in the pool and mounted the platform in the normal ways. With the experimenter blind with respect to genotype, we monitored the animals as they learned to approach a platform marked with a visible cue (Fig. 5a). Both wild-type and PSD-95 mutants reduced the length of their path across the training period (Fig. 5a). Although the mutant mice were initially slower, no difference between groups was seen over the last 6 trials ($P < 0.05$). We then tested spatial learning using the hidden platform version of the watermaze. The mutant mice had significantly longer swim paths than wild-type mice (controls, 4.32 ± 0.30 m; PSD-95 mutants, 9.04 ± 0.50 m; $F(1, 19) = 5.66$, $P < 0.05$; Fig. 5b). After 20 training trials, performance in a transfer test (with the platform removed) was used as an index of learning (Fig. 5c, d). Wild-type mice showed a spatial bias towards the training quadrant, spending significantly more time searching there than in the other three quadrants ($F(3, 24) = 13.84$; $P < 0.0001$), unlike the mutant mice ($F < 1$, $P < 0.9$; group by quadrant interaction, $F(3, 57) = 5.41$; $P < 0.005$). The mice were then given further training using a smaller platform until they reached a specified criterion, or until 32 trials had been completed. All controls (9/9) reached this criterion in 15.1 ± 2.4 trials whereas only 2/12 PSD-95 mutants were successful, giving a group mean of 29.3 ± 2.0 trials ($F(1, 9) = 20.51$; $P < 0.0005$). The PSD-95-mutant mice therefore had a marked inability to learn the position of the hidden platform, an effect also seen when NMDA receptors are blocked^{22,23}.

Discussion

PSD-95 has been suggested in *in vitro* studies⁷ to localize NMDA receptors to the synapse and our results show that they are synaptically localized in PSD-95-mutant mice. This raises the possibility that other molecules may be more important in localizing NMDA-receptor subunits to the synapse. Receptor localization could be achieved by actin in dendritic spines²⁴; actin may interact with NMDA-receptor subunits through α -actinin²⁵ and spectrin²⁶. The NMDA receptor may experience two sets of interactions: one controlling receptor localization and turnover, which is independent of PSD-95, and a second operating through PSD-95 on signal-transduction pathways that control synaptic strength (Fig. 6a).

A wide range of frequencies and patterns of pre- and postsynaptic activity probably act through NMDA-receptor-dependent forms of synaptic plasticity to modulate dynamically the strength of excitatory synaptic transmission *in vivo*. In our PSD-95-mutant mice, the frequency-response function controlling the induction of persistent changes in synaptic efficacy²⁴ showed a dramatic leftward and upward shift so that all frequencies tested (1, 5, 10, 20, 100 Hz) induced LTP. This propensity towards potentiation was most marked at low frequencies, where for example 1-Hz stimulation produced no potentiation in wild-type mice but a large LTP in the PSD-95 mutants. This enhanced LTP requires the activation of NMDA receptors and reaches the enhanced level during the first

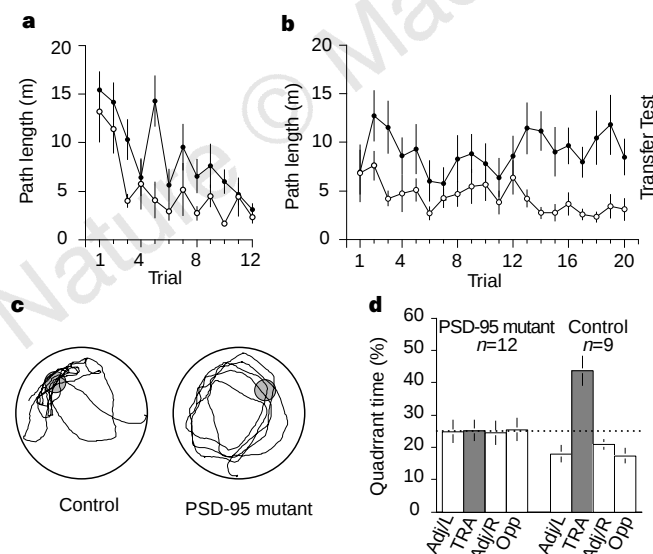


Figure 5 Behavioural data. **a**, Path length across 3 days of training to a visible cue (4 trials per day). By trial 12, the path lengths of both groups in the 2-m-diameter pool were equivalent and close to the minimum possible. Control, white symbols; PSD-95, filled symbols (mean $\pm$ 1 s.e.m.). **b**, Path length across 5 days of place-learning to a hidden platform. Wild-type controls learn to take relatively direct paths, whereas PSD-mutant mice persist in taking circuitous routes. **c**, **d**, Representative swim paths (**c**) and per cent time spent in each quadrant (**d**) during the 60-s transfer test in which the hidden platform (grey in **c**) is not present in the pool. Wild-type mice consistently search in the correct location but PSD-95 mutant mice swim all over the pool. Further training (for up to 32 trials) failed to reveal spatial learning in PSD-95 mutant mice (see text). The horizontal dotted line at 25% in **d** indicates chance performance. Training quadrant, TRA; adjacent left, Adj/L; adjacent right, Adj/R; opposite, Opp.

30 min after high-frequency stimulation (100 Hz). Postsynaptic kinases and phosphatases are known to regulate long-term plasticity during this time window^{1,2}, which suggests that a pathway downstream from the NMDA receptor has been modified in the PSD-95 mutants. Components of this pathway could include PSD-95-binding proteins such as SynGAP^{28,29}, neuroligin¹⁸, GKAP/SAPAP^{30,31}, neuronal NO synthase¹³ and postsynaptic phosphatases that participate in LTD. A model of the NMDA receptor, PSD-95 and associated proteins in a signal-transduction complex (Fig. 6a)

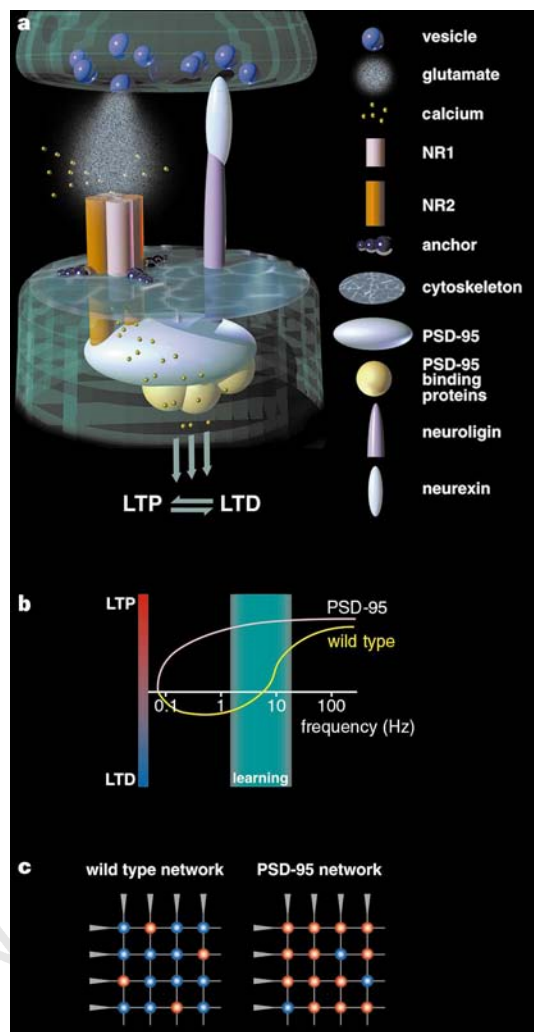


Figure 6 Model for PSD-95 function. **a**, The NMDA-receptor/PSD-95 complex at the synapse with presynaptic terminal with synaptic vesicles (top) and dendritic spine (bottom). NR2 subunits bind PSD-95, which binds neuronal NO synthase, SynGAP, GKAP/SAPAP proteins and neuroligin, which binds the presynaptic protein neuroligin. Anchoring proteins tether the transduction complex through NMDA-receptor subunits to the postsynaptic cytoskeleton. Glutamate release from the presynaptic terminal leads to Ca²⁺ influx, which together with PSD-95 and associated proteins may regulate the balance between LTP and LTD (arrows). **b**, **c**, A model to explain the learning defects in PSD-95 mutant mice, based on learning rules using bidirectional synaptic plasticity. **b**, Postsynaptic activity (as a function of the frequency of trains of synaptic stimulation in Hz) is plotted against the corresponding degree of LTP (red) and LTD (blue). The green box (labelled learning) shows the firing frequency of hippocampal CA1 neurons during exploratory behaviour: in this range, synapses in wild-type mice may show either LTP or LTD, whereas PSD-95 mutant synapses are more likely to show LTP. **c**, This is indicated in a neural network in which PSD-95 mutants have an inappropriately higher number of synapses showing LTP (red) and too few showing LTD (blue). Grey triangles are cell bodies, and processes are lines that intersect at synapses; the colour (blue-red) indicates the strength of synaptic transmission.

in which the associated proteins participate in downstream signalling is in keeping with the 'transducisome' model described for *Drosophila*, where the Trp calcium channel binds PDZ-domain-containing proteins that mediate assembly of a signalling complex³². One function of this complex that is suggested by our paired-pulse facilitation data is to allow the postsynaptic NMDA receptor to regulate transmitter release as a result of PSD-95/neuroligin/neurexin interaction with the presynaptic terminal.

Mathematical models of synaptic plasticity that include bidirectional modifications of synaptic strength (potentiation and depression)^{27,33} when incorporated into neural-network simulations have implications that bear on our PSD-95-mutant mice (Fig. 6b, c). Unlike networks composed of synapses exhibiting only potentiation of synaptic strength, bidirectionally modifiable synapses can increase storage capacity, reduce errors, and limit the number and strength of potentiated synapses to an optimum for memory storage^{34,35}. Hippocampal neurons fire in the θ range of 4–12 Hz during exploratory behaviour³⁶, and in this range neurons in PSD-95 mutants show dramatic LTP, unlike neurons in wild-type mice which are near the threshold between LTP and LTD. Consequently, training of PSD-95-mutant mice in a spatial learning task might cause too many synapses within the network to become strongly potentiated and not enough to be depressed, resulting in a degradation of information storage and recall capacity, which manifests as a learning impairment. Genetic manipulation of PSD-95 will enable bidirectional regulation of synaptic strength to be controlled and provide further insight into synaptic plasticity and neural networks and their integrated function in learning and memory. □

Methods

Gene targeting and biochemistry. The targeting vector comprised 1.6-kb and 4.2-kb bands of, respectively, 5' and 3' genomic DNA flanking a cassette (TAG3IRES1acZpA-MC1neopA) containing the positive selectable marker MC1-neo (neomycin). The 5' end of this cassette contained stop codons in all reading frames to terminate translation of PSD-95 at the *PvuII* site in PDZ3, followed by an internal ribosome-entry site (IRES) which allows a β -galactosidase reporter gene to be expressed under the control of PSD-95. For negative selection, two copies of the thymidine kinase gene (MC1-tk) was ligated 5' to the 5' homology arm. The targeting construct was linearized and electroporated into ES cells (E14Tg2 α IV clone). For Southern blot analysis, 5–15 μ g DNA from ES cells or tail tips were digested with *EcoRI* or *HindIII* and probed with a 0.8-kb cDNA fragment (5' probe) or a 1.2-kb genomic DNA (3' probe) flanking the homology region. Genotyping using PCR required two independent reactions. The wild-type allele was amplified (220-bp product) using a forward primer upstream of the *PvuII* site in the PDZ3 domain (AACCAAGGCGGATCGTGATCCA) and a reverse primer (TCTCTTTGGTGGGCAGTG). The mutant allele (2-kb product) was amplified using a forward primer in the *neo* gene (CATTCGACCACCAAGCG AAGATC) and a reverse primer (CAGGGAGCGGGGACGGATGA) in PSD-95. Amplification was for 30 cycles of 30 s at 93°C, 60 s at 55°C, 60 s at 72°C. Standard procedures were used for synaptosome preparation, protein extraction, immunoblotting and immunoprecipitation. Antibodies: against NR1, mAb 54.1 and pAb 1516 (Chemicon); against PSD-95, pAb138, raised against the N-terminal 438 residues of a mouse PSD-95 fusion protein, and mAb PSD-95 (Transduction Labs).

Neuroanatomy. All procedures relating to the care and treatment of mice conformed with institutional and NIH guidelines. Nine mice were anaesthetized with a lethal dose of chloral hydrate and perfused with aldehydes, dissected, post-fixed, and sectioned³⁷. 50- μ m vibratome sections were used for immunohistochemistry and 15- μ m freshly frozen sections were used for X-gal (5-bromo-4-chloro-3-indolyl- β -galactoside) (Nova Biochem) staining. Primary antibodies were: anti-NR1 mAb 54.1 (1/100); anti-synaptophysin mAb SY38 (1/10; Boehringer); non-phosphorylated neurofilament mAb SMI32 (1/7,500; Steinburger Monoclonal); MAP2 mAb HM2 (1/800; Sigma Immuno Chemicals). Standard procedures were followed for immunofluorescent staining, mounting and coverslipping³⁷. Sections were examined and photographed by experimenters blind to genotype. For post-embedding immunogold localization, six additional mice were perfused with a fixative composed of 2.5% glutaraldehyde, 1% paraformaldehyde and 0.1% picric acid in 0.1M PBS, pH

7.3. After fixation, vibratome sections of CA1 were collected, embedded in resin, thin-sectioned, prepared for post-embedding immunogold localization of NR1, and viewed at 60 kV on a Zeiss CH-10 electron microscope, as described³⁷. Control experiments in which grids were incubated in buffer instead of the NR1 antibody showed no immunogold labelling.

Electrophysiology. 400- μ m-thick slices of mouse hippocampus were maintained at 30°C in an interface-type recording chamber perfused with a murine artificial cerebrospinal fluid (ACSF) containing 124 mM NaCl, 4.4 mM KCl, 25 mM Na₂HCO₃, 1 mM NaH₂PO₄, 1.2 mM MgSO₄, 2 mM CaCl₂ and 10 mM glucose. fEPSPs elicited by Schaffer collateral/commissural fibre stimulation were recorded in the CA1 region as described²⁰. Glass microelectrodes filled with 2 M CsCl (resistance, 60–140 M Ω hms) were used to record EPSPs from individual CA1 pyramidal cells and current injected through the recording electrode was used to hyperpolarize cells to between –90 and –95 mV. There were no differences between mutant and wild-type cells in either resting membrane potential or input resistance. In these experiments, slices with the CA3 region removed were bathed in a modified, 100 μ M picrotoxin-containing ACSF in which the concentrations of CaCl₂ and MgSO₄ were raised to 4.0 mM each and the concentration of KCl was reduced to 2.4 mM. High-frequency-stimulated LTP was induced using 2-s and 1-s trains of 100-Hz stimulation (intertrain interval was 10 s). Saturating levels of LTP were induced by six deliveries of single 1-s trains of 100-Hz stimulation every 5 min. To test synaptic responses mediated by NMDA receptors in mutant and wild-type mice, we used whole-cell voltage-clamp techniques to record EPSCs due to Schaffer collateral fibre stimulation in CA1 pyramidal cells held at membrane potentials between –80 and +40 mV. Low-resistance (3–5 M Ω ; access resistances ranged from 12.5 to 27 M Ω) patch-clamp electrodes were filled with a solution containing 140 mM CsCl, 1 mM MgCl₂, 0.2 mM EGTA, 2 mM Mg-ATP, 0.3 mM GTP and 10 mM HEPES, pH 7.2. In some experiments, electrodes were filled with a solution containing 122.5 mM caesium gluconate, 15 mM CsCl, 2.5 mM TEA-Cl, 1.4 mM NaCl, 0.2 mM EGTA, 2 mM Mg-ATP, 0.3 mM GTP and 10 HEPES, pH 7.2. In these experiments, slices with the CA3 region removed were maintained at room temperature (20 to 22°C) in a submerged recording chamber perfused (2–3 ml per min) with a modified ACSF containing 100 μ M picrotoxin, 2.4 mM KCl, and double the normal concentrations of CaCl₂ and MgSO₄. The AMPA- and NMDA-receptor-mediated components of the EPSCs were estimated from the amplitude of the synaptic currents measured at 5 and 50 ms after the start of the EPSC, respectively. To compare across cells the size of the NMDA-receptor-mediated component of the EPSC, it was normalized to the size of the AMPA-receptor-mediated component. All values are reported as mean $\pm$ s.e.m., *n* being the number of mice. All experiments were performed blind, and after data collection the genotypes were revealed for analysis. Primary cultures of forebrain neurons were prepared from individual neonatal mice (P1) as described³⁸. Whole-cell patch-clamp recordings were made from cells cultured for 12–20 days (wild type, mean age of 13.6 d; PSD-95 mutant, mean age of 16.3 d), which was after the age at which PSD-95 expression was observed by using X-gal staining. The composition of the recording electrode solution was (in mM): CsMeSO₃ 100, Cs-BAPTA 5, HEPES 15, Mg-ATP 4, Na-GTP 0.4, sucrose 40, adjusted to pH 7.2 with CsOH. The composition of the bathing medium was (in mM): NaCl 140, KCl 3, CaCl₂ 2.5, HEPES 15, glucose 10, MgCl₂ 0 or 1, adjusted to pH 7.4 with NaOH. Currents were evoked by brief (50-ms) application of 250 μ M NMDA + 25 μ M glycine. The calculated junction potential of +15 mV was not corrected for in the data shown here. Student *t*-tests were used to assess statistical significance.

Behavioural testing. We used an open-field watermaze (2 m in diameter, opaque water, 25 $\pm$ 1°C, automated swim-path monitoring). Visible platform training: mice were trained to a randomly located platform with a striped flag protruding above it (4 trials per day for 3 days; 30-cm-diameter platform; curtains drawn around the pool to occlude extra-maze cues; maximum trial duration was 90 s; intertrial interval (ITI), 10 min). Hidden platform training: we used a hidden platform with the extra-maze cues visible (4 trials per day, 5 days, 30-cm platform; platform area/pool area was 1/44; 30 s was spent on the platform at the end of each trial; ITI, 10 min). Transfer test: 10 min after the previous training trial, mice were placed in the pool for 60 s (platform absent; the start position was opposite to whatever training quadrant had been used for an individual animal). Training to criterion: extended hidden-platform training, using a smaller hidden platform (20 cm diameter; platform area/pool area

was 1/100) until the animals completed 2 consecutive days with each trial of <20 s, or until 32 trials had been done.

Received 9 June; accepted 13 October 1998.

1. Bear, M. F. & Malenka, R. C. Synaptic plasticity: LTP and LTD. *Curr. Opin. Neurobiol.* **4**, 389–399 (1994).
2. Bliss, T. V. P. & Collingridge, G. L. A synaptic model of memory: long-term potentiation in the hippocampus. *Nature* **361**, 31–39 (1993).
3. Hollman, M. & Heinemann, S. Cloned glutamate receptors. *Annu. Rev. Neurosci.* **17**, 31–108 (1994).
4. Nakanishi, S. & Masu, M. Molecular diversity and functions of glutamate receptors. *Annu. Rev. Biophys. Biomolec. Struct.* **23**, 314–348 (1994).
5. Smart, T. G. Regulation of excitatory and inhibitory neurotransmitter-gated ion channels by protein phosphorylation. *Curr. Opin. Neurobiol.* **7**, 358–367 (1997).
6. Kornau, H.-C., Schenker, L. T., Kennedy, M. B. & Seeburg, P. H. Domain interaction between NMDA receptor subunits and the postsynaptic density protein PSD-95. *Science* **269**, 1737–1740 (1995).
7. Kim, E., Cho, K.-O., Rothschild, A. & Sheng, M. Heteromultimerization and NMDA receptor-clustering activity of chapsyn-110, a member of the PSD-95 family of proteins. *Neuron* **17**, 103–113 (1996).
8. Müller, B. M. et al. SAP102, a novel postsynaptic protein that interacts with NMDA receptor complexes *in vivo*. *Neuron* **17**, 255–265 (1996).
9. Niethammer, M., Kim, E. & Sheng, M. Interaction between the C terminus of NMDA receptor subunits and multiple members of the PSD-95 family of membrane-associated guanylate kinases. *J. Neurosci.* **16**, 2157–2163 (1996).
10. Cho, K. O., Hunt, C. A. & Kennedy, M. B. The rat brain postsynaptic density fraction contains a homolog of the *Drosophila* discs-large tumor suppressor protein. *Neuron* **9**, 929–942 (1992).
11. Kistner, U. et al. SAP90, a rat presynaptic protein related to the product of the *Drosophila* tumor suppressor gene *dlg-A*. *J. Biol. Chem.* **268**, 4580–4583 (1993).
12. Kim, E., Niethammer, M., Rothschild, A., Jan, Y. N. & Sheng, M. Clustering of shaker-type K⁺ channels by interaction with a family of membrane-associated guanylate kinases. *Nature* **378**, 85–88 (1995).
13. Brenman, J. E. et al. Interaction of nitric oxide synthase with the postsynaptic density protein PSD-95 and alpha1-syntrophin mediated by PDZ domains. *Cell* **84**, 757–767 (1996).
14. Hough, C. D., Woods, D. F., Park, S. & Bryant, P. J. Organizing a functional junctional complex requires specific domains of the *Drosophila* MAGUK discs-large. *Genes Dev.* **11**, 3242–3252 (1997).
15. Müller, B. M. et al. Molecular characterization and spatial distribution of SAP97, a novel presynaptic protein homologous to SAP90 and the *Drosophila* discs-large tumor suppressor protein. *J. Neurosci.* **15**, 2354–2366 (1995).
16. Hunt, C. A., Schenker, L. J. & Kennedy, M. B. PSD-95 is associated with the postsynaptic density and not with the presynaptic membrane in forebrain synapses. *J. Neurosci.* **16**, 1380–1388 (1996).
17. Wang, J. H. & Kelly, P. A. T. Attenuation of paired-pulse facilitation associated with synaptic potentiation mediated by postsynaptic mechanisms. *J. Neurophysiol.* **78**, 2707–2716 (1997).
18. Irie, M. et al. Binding of neuroligins to PSD-95. *Science* **277**, 1511–1515 (1997).
19. Tejedor, F. J. et al. Essential role for *dlg* in synaptic clustering of shaker K⁺ channels *in vivo*. *J. Neurosci.* **17**, 152–159 (1997).
20. Mayford, M., Wang, J., Kandel, E. R. & O'Dell, T. J. CaMKII regulates the frequency-response function of hippocampal synapses for the production of both LTD and LTP. *Cell* **81**, 891–904 (1995).
21. Dudek, S. M. & Bear, M. F. Bidirectional long-term modification of synaptic effectiveness in the adult and immature hippocampus. *J. Neurosci.* **13**, 2910–2918 (1993).
22. Morris, R. G. M., Anderson, E., Lynch, G. S. & Baudry, M. Selective impairment of learning and blockade of long-term potentiation by an *N*-methyl-D-aspartate receptor antagonist, AP5. *Nature* **319**, 774–776 (1986).
23. Tsien, J. Z., Huerta, P. T. & Tonegawa, S. The essential role of hippocampal CA1 NMDA receptor-dependent synaptic plasticity in spatial memory. *Cell* **87**, 1327–1338 (1996).
24. Allison, D. W., Gelfand, V. I., Spector, I. & Craig, A. M. Role of actin in anchoring postsynaptic receptors in cultured hippocampal neurons: Differential attachment of NMDA versus AMPA receptors. *J. Neurosci.* **18**, 2423–2436 (1998).
25. Wyszynski, M. et al. Competitive binding of α -actinin and calmodulin to the NMDA receptor. *Nature* **385**, 439–442 (1997).
26. Wechsler, A. & Teichberg, V. I. Brain spectrin binding to the NMDA receptor is regulated by phosphorylation, calcium and calmodulin. *EMBO J.* **17**, 3931–3939 (1998).
27. Bienenstock, E., Cooper, L. & Munro, P. Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *J. Neurosci.* **2**, 32–48 (1982).
28. Kim, J. H., Liao, D., Lau, L.-F. & Huganir, R. SynGAP: a synaptic RasGAP that associates with the PSD-95/SAP90 protein family. *Neuron* **20**, 683–691 (1998).
29. Chen, H.-J., Rojas-Soto, M., Oguni, A. & Kennedy, M. B. A synaptic Ras GTPase-activating protein (p135 SynGAP) inhibited by CaM kinase II. *Neuron* **20**, 895–904 (1998).
30. Naisbitt, S. et al. Characterization of guanylate kinase-associated protein, a postsynaptic density protein at excitatory synapses that interacts directly with postsynaptic density-95/synapse-associated protein 90. *J. Neurosci.* **17**, 5687–5696 (1997).
31. Takeuchi, M. et al. SAPAPs: A family of PSD-95/SAP90-associated proteins localized at postsynaptic density. *J. Biol. Chem.* **272**, 11943–11951 (1997).
32. Tsunoda, S. et al. A multivalent PDZ-domain protein assembles signalling complexes in a G-protein-coupled cascade. *Nature* **388**, 243–249 (1997).
33. Sejnowski, T. J. Statistical constraints on synaptic plasticity. *J. Theor. Biol.* **69**, 387–389 (1977).
34. Willshaw, D. & Dayan, P. Optimal plasticity from matrix memories: what goes up must come down. *Neural Comp.* **2**, 85–93 (1990).
35. Hancock, P. J. B., Smith, L. S. & Phillips, W. A. A biologically supported error-correcting learning rule. *Neural Comp.* **3**, 201–212 (1991).
36. Otto, T., Eichenbaum, H., Wiener, S. I. & Wible, C. G. Learning-related patterns of CA1 spike trains parallel stimulation parameters optimal for inducing hippocampal long-term potentiation. *Hippocampus* **1**, 181–192 (1991).
37. Morrison, B. M., Janssen, W. G. M., Gordon, J. W. & Morrison, J. H. Light and electron microscopic distribution of the AMPA receptor subunit, GluR2, in the spinal cord of control and GluR2 mutant superoxide dismutase transgenic mice. *J. Comp. Neurol.* **395**, 523–534 (1998).
38. Rayport, S. et al. Identified postnatal mesolimbic dopamine neurons in culture: morphology and electrophysiology. *J. Neurosci.* **12**, 4264–4280 (1992).

Acknowledgements. We thank A. J. H. Smith for advice and for the 129 genomic library, J. Ure for embryo injection, L. Anderson and J. Young for mouse care, W. Janssen for technical assistance, F. Johnston and G. Brown for photography, and R. Ellaway for illustrations. This work was supported by the Fyssen Foundation (M.M.) and the BBSRC (M.M., M.D. and P.C.), by grants from the NIH and the Charles A. Dana Foundation (J.H.M. and Y.H.), the Esther A. and Joseph Klingenstein Fund and the Pew Charitable Trusts (T.J.O.), a grant from the National Institute of Mental Health (T.J.O.), and the Wellcome Trust (L.C.W. and S.G.N.G.).

Correspondence and requests for materials should be addressed to S.G.N.G. (e-mail: seth.grant@ed.ac.uk).

Counter-streaming gas flows in solar prominences as evidence for vertical magnetic fields

J. B. Zirker*, O. Engvold† & S. F. Martin‡

* National Solar Observatory, Sunspot, New Mexico 88349, USA

† Institute of Theoretical Astrophysics, University of Oslo, PP Box 1029, Blindern, N0315, Oslo, Norway

‡ Helio Research, 5212 Maryland Avenue, La Crescenta, California 91214, USA

Solar prominences are sheets of relatively cool and dense gas embedded in the surrounding hotter corona. An erupting prominence can inject a mass of up to 10^{15} g into the solar wind¹ as part of a coronal mass ejection. These eruptions must depend critically on the topology of the prominence's magnetic field. In all present models^{2,3}, the prominence hangs on horizontal or helical field lines, while an overlying magnetic arcade temporarily restrains the prominence from erupting. Such models are inconsistent, however, with the slow upward vertical gas flows that are seen in prominences^{4–14}. Here we report counter-streaming flows along closely spaced vertical regions of a prominence, between its top and the lower solar atmosphere. As the flows must be aligned with the magnetic field, this observation implies that a field connects the prominence directly to the photosphere, contrary to all existing models. These magnetic 'tethers' might help prevent a prominence from erupting.

For many years, direct measurements of prominence magnetic fields¹⁵ have been in conflict with observations of vertical flows. The magnetic measurements, which pertain to heights above $\sim 10,000$ km, indicate a horizontal field vector in the plane of the prominence sheet, which modellers take as a necessary condition for static mechanical support of the prominence mass. In models, the mass is supported in dips in a magnetic arcade or in the bottoms of the coils of a horizontal magnetic helix. The helical model¹⁶ is attractive because it accounts both for the existence of the well-known 'cavity' around prominences and for the existence of helical magnetic clouds in interplanetary space following an eruption¹⁷. But the vertical mass flows have been largely ignored and, at first glance, are inconsistent with helical field lines.

To try to resolve this long-standing paradox, we observed motions in the filament shown in Fig. 1 at the Big Bear Solar Observatory for up to 8 hours each day during 14–20 April 1993. We recorded time-lapse images like those in Fig. 1 concurrently in the red and blue sides of the H α line to detect mass motions. Such images in the two wings of H α reveal mass motions towards and away from the observer, as well as transverse motions in the plane of the sky.

The filament has two linked categories of structure: 'spines' and 'barbs'. The spine is the highest, horizontal axial part of the filament, and is composed of resolvable threads. The barbs connect to the high spine and terminate in the chromosphere; they are sheets of mass, composed of parallel narrow threads similar to those in the spine. Figure 1 illustrates that these threads incline steeply between the chromosphere and the spine. The barb threads make a smooth transition into the spine.

Processing of this data revealed (to our knowledge, for the first time) a striking large-scale pattern of mass motions. Simultaneous flows, in opposite directions, are seen in adjacent threads of the barbs as well as in the threads along the spine. Knots of mass can be tracked for distances of 10,000 to 100,000 km at speeds of 5–20 km s⁻¹. In the blue wing on 16 April (Fig. 2 top), the streaming is

only westward along the spine and downward into the barbs on the near side of the filament. In the red wing, the streaming is only eastward along the spine and up in the same barbs on the near side. This systematic pattern appears most clearly on 16 April, when the filament lay just inside the east limb (Fig. 1). Finally, on 20 April, when the filament had crossed the central meridian, the pattern reverses, with a predominant eastward streaming in the blue wing although both eastward and westward streaming are seen in different filament threads.

We designate these concurrent flows in opposite directions as 'counter-streaming'. The counter-streaming threads pervade the entire filament, and must lie so close to each other that the filament has almost the same overall form in the red and blue wings as in Fig. 1, top and bottom. Adjacent threads with opposite motions are sometimes superposed in the line of sight while others have separations of only a few arc seconds. The counter-streaming is especially clear in the spine, but the same phenomenon also occurs in the barbs. Discrete knots flow from the chromosphere into the axial spine while other knots flow in reverse. Individual knots take as long as an hour to rise from the chromosphere to the spine and then continue along the spine.

All of these observations are consistent with steady bidirectional streaming everywhere in the filament along adjacent closely spaced threads. The pattern is observed in both wings of H α , thereby confirming that the flows are mass motions and are not caused by some kind of excitation wave.

The slow drifts of mass (both up and down) along nearly vertical

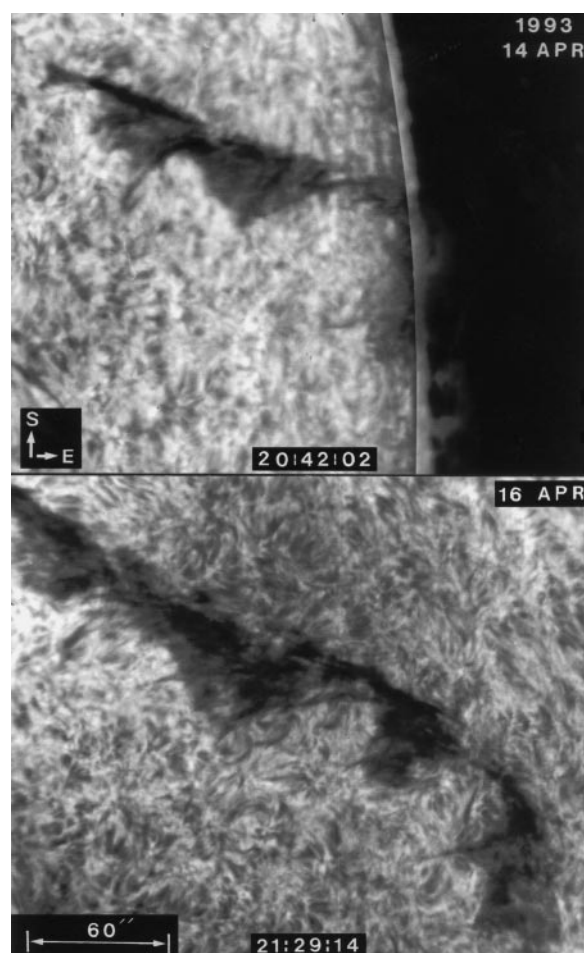


Figure 1 High-resolution H α images of the filament on 14 and 16 April 1993. On 14 April (top) the filter was set at H α -0.3 and H α $+0.3$ Å, with band-passes of 0.5 and 0.25 Å. On 16 April (bottom) the filter was set at H α -0.25 and $+0.25$ Å, with band-passes of 0.5 and 0.25 Å.

field lines are puzzling, however. In principle, the plasma should fall freely along such lines, reaching speeds of 50 km s^{-1} or more, but that conflicts with the speeds in fact seen (a few kilometres per second). A somewhat similar situation prevails in spicules, also without a satisfactory explanation.

Why has counter-streaming not been detected in previous Doppler observations of filaments? We made Dopplergrams from pairs of red-wing and blue-wing images, as have many other authors^{4–14}. We concluded that our line-of-sight velocity maps alone do not reveal counter-streaming because a simple subtraction of moderate-resolution images, taken in opposite wings, will usually result in a null Doppler signal. Second, motions of the chromospheric background can corrupt the signal in the filament, unless spatial resolution is sufficiently high. Last, isolated pairs of images do not reveal the long distances traversed by the moving knots of mass.

Detection of the counter-streaming requires high spatial resolution, a favourable filament orientation, time-series of observations in both wings simultaneously over several hours, carefully registered images and a high rate of image projection. The filament we observed is quite typical. We therefore suggest that counter-streaming is a feature common to all quiescent filaments.

Our observations contradict all contemporary prominence models, even those of barbs, as these models do not allow any

vertical motions, let alone counter-streaming. Simply allowing horizontal field lines to drift up and down, while carrying blobs of mass, will not easily explain counter-streaming, for then alternate field lines would have to pass each other and travel long diagonal distances. Such complex motions seem unrealistic to us.

We propose instead a conceptual model¹⁸ in which field lines bend down from the spine to the chromosphere (see Fig. 2 centre). In this model, the mass flows along the field lines: it is not carried by moving horizontal field lines, and does not cross the field lines. The tension of such steeply inclined field lines, which are embedded in the prominence sheet, might act to restrain the prominence from erupting. Certainly the existence of such field lines needs to be incorporated in any acceptable model of prominence eruption. □

Received 24 April; accepted 10 September 1998.

1. Jackson, B. V. & Howard, R. CME mass distribution derived from SOLWIND coronagraph observations. *Sol. Phys.* **148**, 359–370 (1993).
2. Priest, E., McKay, D. & Longbottom, A. Dipped magnetic field configuration associated with filaments and barbs. *Astron. Astrophys.* (in the press).
3. Aulanier, G., Demoulin, P., van Driel-Gesztely, L., Mein, P. & DeForest, C. 3-D magnetic configurations supporting prominences II. *Astron. Astrophys.* **335**, 309–322 (1998).
4. Dunn, R. B. *A Photometric Investigation of the Solar Chromosphere*. Thesis, Harvard Univ. (1960).
5. Priest, E. O. The small-scale velocity field of a quiescent prominence. *Sol. Phys.* **70**, 315 (1981).
6. Engvold, O. in *New Perspectives on Solar Prominences* (eds Webb, D., Rust, D. M. & Schneider, B.) (IAU Colloq. 167, Kluwer Academic Press, in the press).
7. Engvold, O. The fine structure of prominences I. *Sol. Phys.* **49**, 283–295 (1976).
8. Liggett, M. & Zirin, H. Rotation in prominences. *Sol. Phys.* **91**, 259–267 (1984).
9. Mein, P. in *Solar Coronal Structures* (eds Rusin, V., Heinzel, P. & Vial, J. C.) 289–296 (IAU Colloq. 144, Veda Publishing House, 1993).
10. Martres, M. J., Mein, P., Schneider, B. & Soru-Escout, I. Structure and evolution in quiescent prominences. *Sol. Phys.* **69**, 301–312 (1981).
11. Mahlerbe, J., Schneider, B. & Mein, P. Dynamics in the filaments. *Astron. Astrophys.* **102**, 124–128 (1981).
12. Mahlerbe, J. M., Schneider, B., Ribes, E. & Mein, P. Mass motions in filaments. *Astron. Astrophys.* **119**, 197–206 (1983).
13. Engvold, O. & Keil, S. L. in *Coronal and Prominence Plasmas* (ed. Poland, A. I.) (NASA Conf. Publ. 2442, NASA, Washington, DC, 1986).
14. Kubota, J. & Uesegi, A. Vertical motion of a prominence. *Publ. Astron. Soc. Jpn* **38**, 903–909 (1986).
15. Leroy, J.-L., Bommier, V. & Sahal-Brechot, S. New data on the magnetic structure of quiescent prominences. *Astron. Astrophys.* **131**, 33–44 (1984).
16. Hundhausen, J. R. & Low, B. C. Magnetostatic structure of the solar corona 2, the magnetic topology of quiescent prominences. *Astrophys. J.* **443**, 818–832 (1995).
17. Bothmer, V. & Schwenn, R. Eruptive prominences as sources of magnetic clouds in the solar wind. *Space Sci. Rev.* **70**, 215–220 (1994).
18. Martin, S. F. & Echols, C. R. in *Solar Surface Magnetism* (eds Rutten, R. J. & Schrijver, C. J.) 339–346 (Kluwer Academic, Dordrecht, 1994).

Acknowledgements. We thank H. Zirin for the observing run at BBSO, D. Martin for reproducing the original photographic data in digital movie format, and R. Ewald for her assistance in preparing the illustrations.

Correspondence and requests for materials should be addressed to J.B.Z. (e-mail: jzirker@noao.edu).

Temperature-induced magnetization reversal in a YVO_3 single crystal

Y. Ren*, T. T. M. Palstra†, D. I. Khomskii*, E. Pellegrin*, A. A. Nugroho‡§, A. A. Menovsky‡ & G. A. Sawatzky*

* Solid State Physics Laboratory, † Inorganic Solid State Chemistry Laboratory, Materials Science Centre, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

‡ Van der Waals-Zeeman Institute, University of Amsterdam, Valckenierstraat 65, 1018 XE Amsterdam, The Netherlands

The total energy of a magnet in a magnetic field is lowest when the magnetic moment is aligned parallel to the magnetic field. Once aligned, the magnetic moment can be reversed by applying a sufficiently large field in the opposite direction. These properties form the basis of most magnetic recording and storage devices. But the phenomenon of magnetization reversal in response to a change in temperature (in a small magnetic field) is rarer. This effect occurs in some ferrimagnetic materials consisting of two or

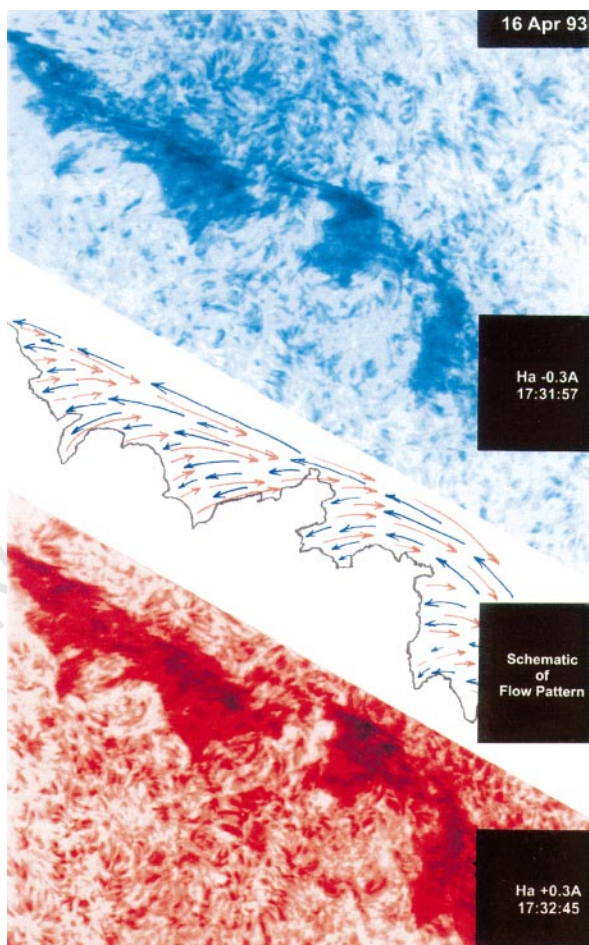


Figure 2 Images on 16 April, showing the inclined threads. In the blue wing, the streaming is only westward along the spine and downward into the barbs on the near side of the filament (blue arrows in the middle panel). In the red wing, the streaming is only eastward along the spine and up in the same barbs on the near side (red arrows in middle panel). This systematic pattern appears most clearly on 16 April, when the filament lay just inside the east limb. Finally, on April 20, when the filament had crossed the central meridian, the pattern reverses, with a predominant eastward streaming in the blue wing although both eastward and westward streaming are seen in different filament threads.

more types of antiferromagnetically ordered magnetic ions¹, and forms the operational basis of ferrimagnetic insulators. Here we report the observation of multiple temperature-induced magnetization reversals in YVO₃. The net magnetic moment is caused by a tilting of the antiferromagnetically aligned moments of (crystallographically identical) V³⁺ ions, due to orthorhombic distortion in the crystal structure. We observe an abrupt switching at 77 K associated with a first-order structural phase transition, and a gradual reversal at ~95 K without an accompanying structural change. The magnetization always reverses if the crystal is cooled or warmed through these two temperatures in modest fields. We propose a possible mechanism involving a change in orbital ordering which may be generic to a broad class of transition metal oxides.

Transition metal oxides with the perovskite structure display a large variety of properties such as high-temperature superconductivity, colossal magnetoresistance^{2,3} and very diverse magnetic properties. We now report on another novel and peculiar phenomenon in YVO₃: multiple and reversible sign changes in the magnetization with temperature. Such magnetic-moment reversals have been observed in those ferrimagnets¹ with strong magnetic anisotropy that exhibit a compensation temperature. The net magnetization, initially oriented parallel to the field, changes sign at this temperature and the metastable, energetically unfavourable state is fixed by the strong anisotropy. For such an effect to occur, inequivalent sites must exist. In YVO₃, however, all magnetic V sites are equivalent, as shown by the crystallographic data.

Close to our situation is the observation of a 'diamagnetic' response in LaVO₃ (refs 4–6). Upon weak-field cooling, this system exhibits a magnetization opposite to the applied magnetic field below a structural phase transition temperature $T_1 = 138 \text{ K} < T_N = 142 \text{ K}$, where T_N is the magnetic ordering (Néel) temperature. It has been suggested that this diamagnetic

response is due to a reversal of a canted-spin moment on traversing the first-order Jahn–Teller phase transition at T_b , below which the orbital angular momentum is maximized^{6–8}, and that the response of the orbital moment to the forces generated at the first-order phase transition can reverse the Dzyaloshinsky vector so as to create a canted spin in a direction opposite to the applied field, given that T_1 is close to T_N . Other polycrystalline samples of V³⁺ perovskites show weak indications of anomalous behaviour^{7–9}. We demonstrate the dramatic behaviour of single crystals of YVO₃ in which multiple magnetization switching is clearly observed. In contrast with LaVO₃, YVO₃ provides a wide temperature window between T_N and the first-order phase transition which we refer to as T_s , enabling us to study the effect in detail. Furthermore, an additional magnetization reversal is observed, irrespective of the cooling conditions.

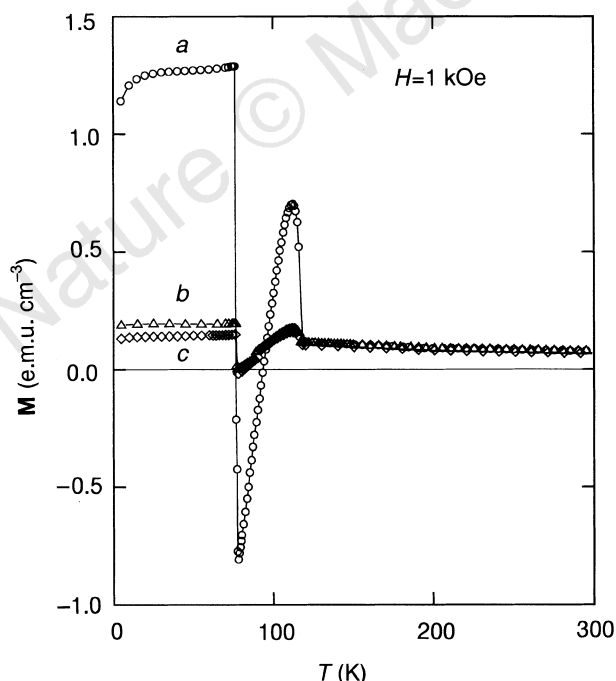


Figure 1 Temperature dependence of the magnetization in an applied magnetic field of 1 kOe along the *a*-, *b*- and *c*-axes, respectively. Upon cooling the sample in a fixed magnetic field $H < 4 \text{ kOe}$ below $T_N = 116 \text{ K}$, the magnetization after first increasing starts decreasing and crosses zero at $T^* \approx 95 \text{ K}$ to a large negative value. With further cooling, it jumps at $T_s = 77 \text{ K}$ to a large positive value. This second transition exhibits hysteresis that is consistent with a first-order transition, as inferred from the structural data. In contrast with LaVO₃, we observe not only an abrupt magnetization reversal at the first-order transition T_b , but also a gradual reversal near 95 K.

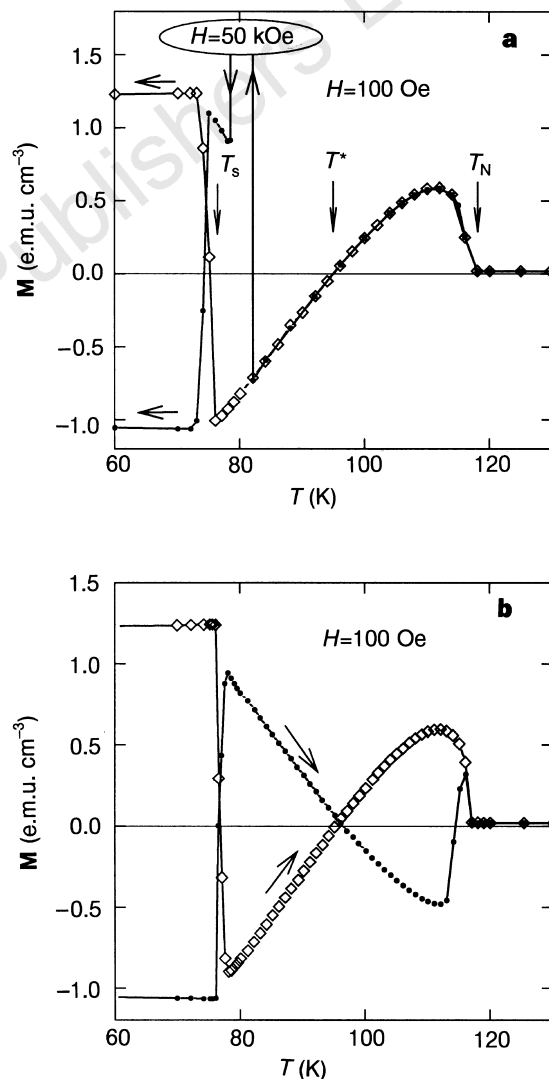


Figure 2 Magnetization versus temperature in a field of 100 Oe, demonstrating the memory effect upon the application of large fields. In the curve marked by filled circles in **a** we follow M with decreasing temperature down to a temperature below T^* ; a high field is then applied to flip the magnetization positive, after which the field is lowered again to 100 Oe and the temperature decreases. The curve marked by diamonds is measured without intermediate application of the high field. In **b**, we show the curves with increasing temperature starting from below T_s : the curve marked by diamonds without having 'trained' the sample, and the curve marked by filled circles after 'training' as described for **a**. This demonstrates the reversibility: upon warming, M now switches from negative below T_s to positive for $T_s < T < T^*$, and becomes negative for $T > T^*$. It is thus nearly a 'mirror image' of the behaviour shown in Fig. 1 and the diamond marked curve in **b**, except that it becomes positive again close to T_N .

YVO_3 has a distorted perovskite structure of the type GdFeO_3 (crystal symmetry $Pbnm$) at all temperatures. It has two magnetic phases: one at $T_s \approx 77 \text{ K} < T < T_N = 116 \text{ K}$ and another at $T < T_s$. The detailed magnetic structure of these phases is controversial^{10,11} but is not crucial for the experimental observations reported here.

The distorted crystal structure naturally leads to canted spin structures. As the oxygen ions mediating the superexchange interaction between the two nearest neighbour V ions are not at an inversion centre, the antisymmetric Dzyaloshinsky–Moriya (DM) interaction of the form $\mathbf{D} \cdot (\mathbf{S}_1 \times \mathbf{S}_2)$ will be present, where $\mathbf{D}$ is the so-called Dzyaloshinsky vector and $\mathbf{S}_1$ ($\mathbf{S}_2$) are the V spins corresponding to the two magnetic sublattices. The DM interaction, which prefers canted spin arrangements, is in competition with the quite different ordinary Heisenberg exchange $\mathbf{J}\mathbf{S}_1 \cdot \mathbf{S}_2$, which prefers collinear spin arrangements. Also, because the oxygen octahedra coordinating the V ions are twisted to form a staggered V–O bond direction along the c -axis, the single-ion anisotropy easy axis is staggered. Both of these mechanisms can lead to weak ferromagnetism¹². The fact that YVO_3 is a weak ferromagnet with spin canting has been established from measurements on powder samples¹³. Several batches of crystallographically pure large single crystals ($5 \times 5 \times 5 \text{ mm}^3$) were grown by the floating-zone method. The magnetic measurements made on a large number of our single crystals all reveal the same unusual behaviour, with several magnetization reversals upon cooling in a weak magnetic field (Fig. 1).

Even more striking, is the observation of a memory effect shown in Fig. 2a. This indicates that the sign of the magnetization can be reversed by the application of a large enough field, but that upon lowering the field, the temperature-dependent net magnetic moment $\mathbf{M}(T)$ always changes sign, irrespective of what its sign was, when crossing T_s both on cooling and warming. The same is also true for the second crossing at T^* .

Although we do not as yet have a detailed microscopic theory for the magnetization reversal with temperature, we propose the model shown in Fig. 3. As mentioned above, there are two mechanisms for producing a canted spin structure in these materials: single-ion magnetic anisotropy and DM coupling. These two canting mechanisms produce the net moment in the a – c plane, and in our model we assume that they are oriented in opposite directions. We checked that the observed phenomena are connected to spin canting: the differential susceptibility $d\mathbf{M}/dH$ is always positive (even in the ‘diamagnetic’ state) and the net moment is ~ 0.01 Bohr magneton per V ion, which for an $S = 1$ system corresponds to a canting angle of $\sim 0.2^\circ$.

Using this model, we present in Fig. 3 a pictorial view of what could happen as a function of temperature. We start at temperatures just below T_N (Fig. 3a). Here, the two sublattice spins prefer to lie close to a local easy axis if the local magnetic anisotropy is large, resulting in a net magnetic moment, as shown. As T decreases, this net moment will first grow because of the development of a sublattice magnetization due to superexchange. However, as the sublattice magnetization develops, so does the DM coupling, which tends to cant the spins in the opposite direction. Consequently, the net moment will reach a maximum and then decrease. It crosses zero at the temperature below which the DM interaction dominates. This will result in a moment opposite to the small applied field. It could only reverse to its lowest energy state in the field by reversing the two sublattices on a macroscopic scale, resulting in a frozen in metastable state (Fig. 3b). A large external magnetic field can overcome the barrier for rotation of the sublattice spins, resulting in a reversal of the sublattice spin orientation (Fig. 3c). The net moment is now oriented parallel to the field and will remain so upon lowering the field. If we now increase the temperature, we return to the situation where magnetic anisotropy dominates. Thus, the net moment will change sign again, now turning negative for $T > 95 \text{ K}$ (the curve marked by filled circles in Figs 2b, 3d). Once we get close enough to T_N , the energy barrier for the reversal of the sublattice magnetization will eventually become very small. Any small field will flip the net moment again to a positive value and reach a maximum just below T_N , dropping to zero at T_N (Fig. 2). This model explains all the data from high temperatures down to the first-order phase transition at 77 K .

The magnetization reversal at T_s remains more of a puzzle. It follows from our results that the ferromagnetic moment is oriented along the a -axis both above and below T_s . From neutron scattering, we know that the magnetic structure changes at T_s from C - to G -type, as proposed by Kawano *et al.*¹¹ This is definitely connected with a change of orbital ordering occurring at T_s . Low-temperature single-crystal X-ray diffraction (Y.R. *et al.*, unpublished results) of YVO_3 shows that, whereas above T_s the V–O bond lengths are almost equal, the first-order phase transition is accompanied by a strong distortion of these bond lengths, resulting in pairs of long ($2.052 \text{ \AA}$), intermediate ($1.992 \text{ \AA}$) and short ($1.975 \text{ \AA}$) V–O bonds. The long and short bonds are oriented alternately along the $[110]$ and $[1\bar{1}0]$ in the basal a – b plane, whereas the intermediate one of $1.992 \text{ \AA}$ is along the c -axis, similar to the structure of LaMnO_3 ; such a distortion corresponds to orbital ordering, with the d_{xy}

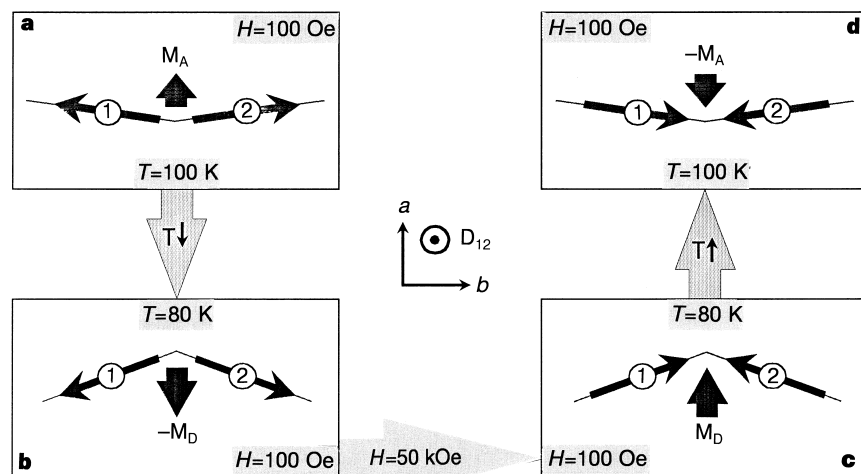


Figure 3 Pictorial view of the temperature dependence of the magnetization. **a**, Just below T_N , the two sublattice spins prefer to lie close to a local easy axis if the local magnetic anisotropy is large, resulting in a net magnetic moment parallel to the applied field. **b**, As T decreases below $T^* = 95 \text{ K}$, the DM interaction dominates which tends to cant the spins in opposite direction. A large external magnetic field

can overcome the barrier for rotation of the sublattice spins, resulting in a reversal of the sublattice spin orientation. **c**, The net moment is now oriented parallel to the field and will remain so upon lowering the field. If we now increase the temperature, we come back to the situation where the magnetic anisotropy dominates. **d**, Thus, the net moment will change sign again, turning negative for $T > 95 \text{ K}$.

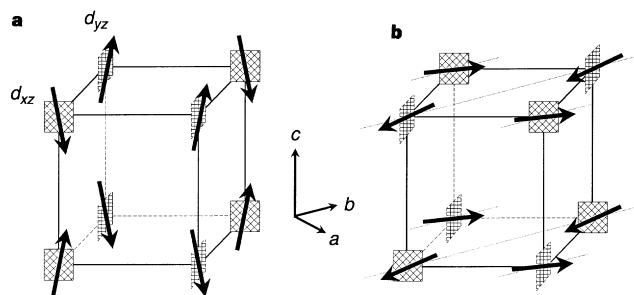


Figure 4 Suggested orbital ordering below and above $T_s = 77$ K. **a**, A C-type orbital ordering with a G-type spin ordering; and **b**, a G-type orbital ordering with a C-type spin ordering. The dotted lines are along the b -axis. The hatched squares indicate the planes in which the occupied d orbitals lie. The d_{xy} orbitals that are in each case occupied by one electron are omitted for clarity.

orbital occupied at each V^{3+} site and the second electron occupying, respectively, the d_{xz} and d_{yz} orbitals in two sublattices as shown in Fig. 4. According to the Goodenough–Kanamori rules¹⁴, this orbital occupation naturally leads to the G-type antiferromagnetism¹¹. A similar result was also obtained by band-structure calculation¹⁵. From Fig. 1, we see that the ferromagnetic component is both above and below T_s oriented parallel to the a -axis of the crystal; the magnetic structure above T_s must therefore be of C-type. Then, according to Bertaut's symmetry considerations¹², the easy axis below T_s should be close to the c direction in order for the weak moment to remain along a . From this, we conclude that the easy axis must indeed change on going through the phase transition. The C-type magnetic ordering observed between T_s and T_N (ref. 11) should then correspond to an orbital structure, with the alternation of the sublattices also being along the c direction and, as discussed above, with the easy axis almost parallel to b . Such a drastic change of the magnetic properties at T_s can lead to a change in the relative role of the DM interaction and the anisotropy. Careful neutron scattering experiments are needed to clarify the behaviour of YVO_3 across T_s .

Our results make it difficult to compare $LaVO_3$ and YVO_3 in detail: single crystals of $LaVO_3$ are needed, although the close proximity of T_N and T_f in $LaVO_3$ will be a complicating factor. $\square$

Received 19 June; accepted 19 August 1998.

1. Menyuk, N., Dwight, K. & Wickham, D. G. Magnetisation reversal and asymmetry in cobalt vanadate (IV). *Phys. Rev. Lett.* **4**, 119–120 (1960).
2. Goodenough, J. B. Metallic oxides. *Prog. Solid State Chem.* **5**, 145–399 (1973).
3. Khomskii, D. I. & Sawatzky, G. A. Interplay between spin, charge and orbital degrees of freedom in magnetic oxides. *Solid State Commun.* **102**, 87–99 (1997).
4. Shirakawa, N. & Ishikawa, M. Anomalous diamagnetism of a perovskite $LaVO_3$. *Jpn Appl. Phys.* **30**, L755–L756 (1991).
5. Mahajan, A. V., Johnston, D. C., Torgeson, D. R. & Borsa, F. Magnetic properties of $LaVO_3$. *Phys. Rev. B* **46**, 10966–10972 (1992).
6. Goodenough, J. B. & Nguyen, H. C. The peculiar ferromagnetism of $LaVO_3$. *C.R. Acad. Sci., Paris* **319**, 1285–1291 (1994).
7. Nguyen, H. C. & Goodenough, J. B. Magnetic studies of some orthovanadates. *Phys. Rev. B* **52**, 324–334 (1995).
8. Nguyen, H. & Goodenough, J. B. Magnetic and transport properties of $CeVO_3$. *J. Solid State Chem.* **119**, 24–35 (1995).
9. Corti, M., Cintolesi, A., Lascialfari, A., Rigamonti, A. & Rossetti, G. Magnetic properties of YVO_3 from susceptibility and ^{89}Y NMR measurements. *J. Appl. Phys.* **81**, 5286–5288 (1997).
10. Zubkov, V. G., Bazuev, G. V. & Shveikin, G. P. Low-temperature neutron- and X-ray-diffraction investigations of rare-earth orthovanadates. *Sov. Phys. Solid State* **18**, 1165–1166 (1976).
11. Kawano, H., Yoshizawa, H. & Ueda, Y. Magnetic behavior of a Mott-insulator YVO_3 . *J. Phys. Soc. Jpn* **63**, 2857–2861 (1995).
12. Bertaut, E. F. in *Magnetism Vol. III* (eds Rado, G. T. & Suhl, H.) 86–126 (Academic, New York and London, 1963).
13. Borukhovich, A. S., Bazuev, G. V. & Shveikin, G. P. Anomalies of the specific heat and magnetic thermodynamic parameters of yttrium orthovanadate. *Sov. Phys. Solid State* **16**, 191–192 (1974).
14. Goodenough, J. B. *Magnetism and the Chemical bond* (Interscience, New York–London, 1963).
15. Sawada, H., Hamada, N., Terakura, K. & Asada, T. Orbital and spin orderings in YVO_3 and $LaVO_3$ in the general gradient approximation. *Phys. Rev. B* **53**, 12742–12749 (1996).

Acknowledgements. We thank J. B. Goodenough for discussion of the results and for explaining his ideas on the origin of the magnetization reversal in $LaVO_3$; we also thank J. Rodriguez-Carvajal and E. F. Bertaut for discussion. This work is supported by the Netherlands Foundation for the Fundamental Research of Matter (FOM) and the Dutch Organization for the Advancement of Pure Research (NWO), and in part by EU (OXSEN).

Correspondence and requests for materials should be addressed to G.A.S.

Spontaneous ordering of bimodal ensembles of nanoscopic gold clusters

C. J. Kiely*, J. Fink*†, M. Brust†, D. Bethell† & D. J. Schiffrin†

* Materials Science and Engineering, Department of Engineering, The University of Liverpool, Liverpool L69 3BX, UK

† Department of Chemistry, The University of Liverpool, Liverpool L69 7ZD, UK

The controlled fabrication of very small structures at scales beyond the current limits of lithographic techniques is a technological goal of great practical and fundamental interest. Important progress has been made over the past few years in the preparation of ordered ensembles of metal and semiconductor nanocrystals^{1–7}. For example, monodisperse fractions of thiol-stabilized gold nanoparticles⁸ have been crystallized into two- and three-dimensional superlattices⁵. Metal particles stabilized by

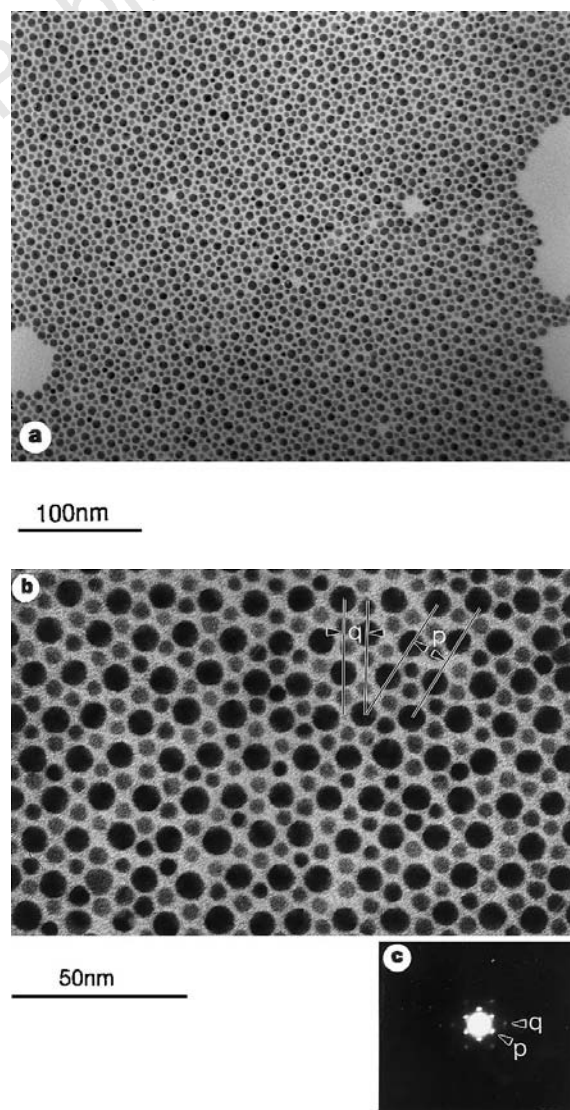


Figure 1 An ordered raft comprising Au nanoparticles of two distinct sizes with $R_B/R_A \approx 0.58$. Shown are electron micrographs at low (**a**) and higher (**b**) magnification. **c**, The low-angle superlattice electron diffraction pattern obtained from this bimodal raft structure.

quaternary ammonium salts can also self-assemble into superlattice structures^{9,10}. Gold particle preparations with quite broad (polydisperse) size distributions also show some tendency to form ordered structures by a process involving spontaneous size segregation^{11,12}. Here we report that alkanethiol-derivatized gold nanocrystals of different, well defined sizes organize themselves spontaneously into complex, ordered two-dimensional arrays that are structurally related to both colloidal crystals and alloys between metals of different atomic radii. We observe three types of organization: first, different-sized particles intimately mixed, forming an ordered bimodal array (Fig. 1); second, size-segregated regions, each containing hexagonal-close-packed monodisperse particles (Fig. 2); and third, a structure in which particles of several different sizes occupy random positions in a pseudo-hexagonal lattice (Fig. 3).

Figure 1a shows an example of an extended area where particles of 4.5 ± 0.8 and 7.8 ± 0.9 nm diameter are packed in a highly organized manner, exhibiting a remarkable degree of long-range order. Here, as in all other cases, adjacent Au clusters were separated by a region of approximately 1 nm which did not exhibit any diffraction contrast. This distance is considerably less than twice the expected C₁₀-thiol chain length; interdigitation of the alkane chains from nearest-neighbour Au particles can be inferred^{13–16}. Figure 1b shows more clearly that the larger particles form an hexagonal array and (in projection) the smaller particles occupy all the trigonal interstices. The lateral extent of these bimodal rafts was typically $1 \mu\text{m}^2$ and approximately ten areas with this particular structure could be found on any transmission electron microscopy (TEM) grid examined. High-resolution electron micrographs (not shown) indicate that the larger particles always display {111} surface facets and internal twinning defects that are characteristic of multiply twinned icosahedral or decahedral particles. The smaller particles, on the other hand, do not exhibit internal twinning suggesting that they are truncated cuboctahedra¹⁷. If the large and small particles are designated A and B respectively, the local stoichiometry in these bimodal regions is exactly AB₂.

Electron diffraction by these rafts gave rise to polycrystalline ring patterns which can be indexed to the usual 111, 200, 220..... reflections of face-centred cubic (f.c.c.) gold. A low-angle spot pattern was observed close to the transmitted beam (Fig. 1c) demonstrating the long-range order of these structures. The innermost hexagonal ring of reflections yields a periodicity of 13 nm, whereas the weaker second hexagonal ring (which is rotated by 30° with respect to the first) corresponds to a periodicity of 7.8 nm. These spatial frequencies correspond well with the superlattice planes marked "p" and "q" in Fig. 1b, which have a measured

interplanar angle of 30° and average periodicities of 13 and 8 nm, respectively. This bimodal ordering contrasts with earlier reports of two-dimensional organization of polydisperse nanoparticles into rafts with an apparent radial distribution of particle sizes, either decreasing¹¹ or increasing¹² with distance from the centre of the area.

There is a clear parallelism between the bimodal structures observed in the present work and those found on the micrometre scale for colloidal crystals, on the one hand, and on the atomic scale in ordered intermetallic alloy structures, on the other. Mixtures of large (A) and small (B) spheres with a radius ratio $R_B/R_A \approx 0.58$ often exhibit an AB₂ arrangement, isostructural, for instance, with the alloy AlB₂ (ref. 18). Similar structures to that shown in Fig. 1 have been observed in Brazilian gem opals^{19,20}, charge-stabilized polystyrene spheres²¹, latex²² and PMMA spheres²³. Using simple geometrical arguments, Murray and Sanders²⁴ suggested that the AB₂ colloidal crystals should only be stable in the ranges $0.482 < R_B/R_A < 0.624$, and phase-diagram calculations^{23,25} have shown that the AB₂ structure is particularly favoured when $R_B/R_A \approx 0.58$. This ratio was indeed observed for the bimodal rafts of Au particles studied here (Fig. 1), showing their similarity to both colloidal crystals and binary alloys.

The projected bimodal raft structure is very similar to the [0001] projection of the AB₂ structure, in which small spheres form planar hexagonal rings which alternate with close-packed layers of the larger spheres, such that the latter are positioned above and below the centre of the B rings. Other alloy structures, however, can produce similar projections. For instance, the basal layer of AB₅ (CaCu₅) has an A:B ratio of 1:2 in which the hexagonal ring structures of the B spheres are coplanar with the hexagonal raft of larger A spheres²¹. The TEM images cannot distinguish whether or not the A and B particles are coplanar, and so it is not possible at present to assign unambiguously an AB₂ or AB₅ type alloy structure to the bimodal raft. Further, depending on the atomic fractions of A and B present, an R_B/R_A value of ~ 0.58 can also result in an AB₁₃ (refs 23, 25) structure, but this has not been observed in the present work. Although the structural details need additional investigation, the results presented here represent an ordered bimodal structure in which the constituent components are on the nanometre scale, thus bridging the gap between the atomic (alloys) and the micrometre/submicrometre (colloidal crystals) length scales.

Figure 2 shows a different type of organization of an approximately bimodal ensemble of particles for which $R_B/R_A \approx 0.47$. The particles within the raft show a distinct tendency to phase-separate into domains of small and large spheres, each domain exhibiting

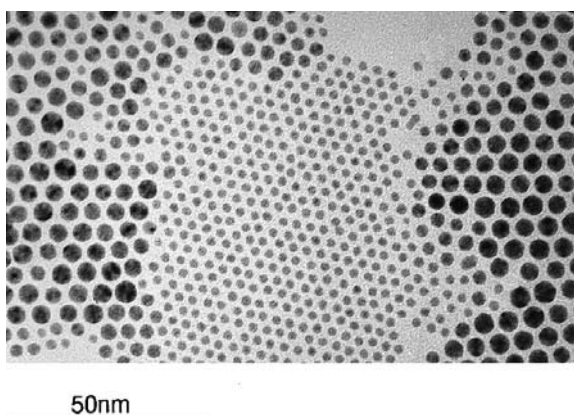


Figure 2 Electron micrograph of a phase-separated A+B mixture of Au nanoparticles obtained when the R_B/R_A ratio is ~ 0.47 . In this case, $R_A = 4.5 \pm 0.7$ nm and $R_B = 9.6 \pm 1.5$ nm.

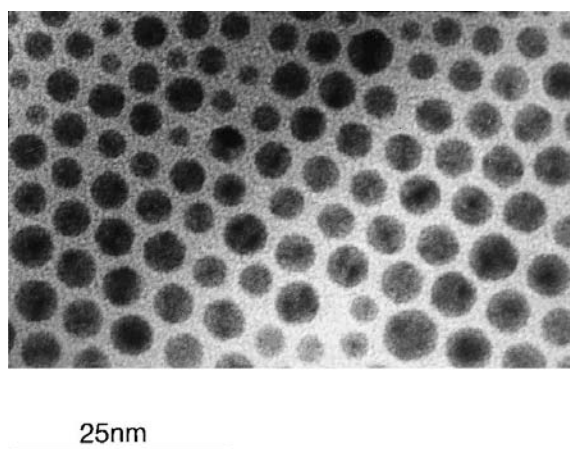


Figure 3 Electron micrograph of a 'random alloy' of Au nanoparticles obtained for an R_B/R_A ratio greater than 0.85.

two-dimensional hexagonal packing. This observation is again in agreement with the work of Murray and Sanders²⁴ who suggested that A and B segregation into two close-packed monodisperse phases should occur for $0.458 < R_B/R_A < 0.482$. In contrast, if R_B and R_A are very similar, as shown in Fig. 3 for $R_B/R_A \approx 0.87$, the structure formed approximates to an intimate random alloy. The latter observation is in accordance with the 15% Hume–Rothery rule for simple alloy theory²⁶ which states that substitutional solid solutions can only be formed when the individual components do not differ by more than 15% in their size.

The coexistence of several types of spatially discrete monodisperse and bidisperse rafts, as repeatedly found in the TEM specimens studied, is highly likely for entropy-driven crystallization²⁵ of a system in which the solution contains a small number of different, but well defined, particle sizes²⁴. The formation of “magic number” sized clusters from the initial preparations by Brust *et al.*⁸ is well documented for cluster sizes up to 3 nm (ref. 27). The success of the preparative method for the thiol-derivatized colloidal solutions described here in obtaining larger clusters of discrete sizes may be related to etching of the ‘naked’ clusters by alkanethiol. Preparations in the presence of thiols as described, for example, by Murray and co-workers lead to particles in the size range 2–6 nm depending on the reaction conditions²⁸.

A final result relates to the long-term stability of these superlattices. The thiol-derivatized Au clusters are stable in toluene for periods of over 3 years (ref. 8). When the particles are allowed to crystallize, the protective alkanethiol coating can slowly degrade under ambient conditions and it becomes energetically favourable for the Au particles to minimize their surface energy by sintering. Figure 4 shows an ordered bimodal raft of the type previously shown in Fig. 1b after exposure to air for several months. The 7.8-nm-diameter particles are still intact, whereas the 4.5-nm particles have spread to wet the carbon substrate. The smaller particles have effectively connected to form a semi-continuous network in which the regular isolated thiol-capped larger particles are embedded. The reason for the decreased stability of smaller particles as compared to the larger ones may be related to their structure. It was previously noted that the larger particles were multiply twinned exposing only {111} surface planes, whereas the smaller particles are truncated cuboctahedra exposing both {111} and {100} surface facets. A further contributing factor is the strong dependence of sintering rate on particle radius. All other conditions (such as temperature, ambient humidity) being equal, the smaller particles are expected to sinter at a relative rate of approximately $(R_A/R_B)^4$; in the present case, this is about 9 times faster than the larger particles²⁹. This raises

the possibility that the size-dependence of particle stability might be exploited for nanopatterning; for example, in source and drain formation in single-electron transistor structures (see ref. 30, and references therein).

The observations described here indicate the possibility of manipulating the composition and structure of nanoscale materials by simply adjusting the size ratios and relative proportions of their constituents (for example, metallic, magnetic or semiconducting particles.) □

Methods

Thiol-capped gold nanoparticles were prepared by the addition of decanethiol to a toluene solution of colloidal gold prepared by established two-phase methods^{3,8}. The conditions of synthesis were slightly modified as follows: after phase transfer of the gold salt, the toluene phase contained 1 mmol of AuCl_4^- in 80 ml of 50 mM tetrakis(decyl)ammonium bromide. After reaction, 1-decanethiol (523 mg) was added; isolation of the derivatized clusters was achieved by precipitation following addition of 200 ml of methanol, and the pure product was redissolved in toluene. Specimens for examination by electron microscopy were prepared by the slow evaporation of one drop of this toluene solution onto a carbon-coated copper mesh grid. All samples were examined in a JEOL 2000EX TEM operating at 200 kV. The structural features described were always present in repeat preparations.

Received 22 June; accepted 24 August 1998.

- Giersgi, M. & Mulvaney, P. Preparation of ordered colloid monolayers by electrophoretic deposition. *Langmuir* **9**, 3408–3413 (1993).
- Mirkin, C. A., Letsinger, R. L., Mucic, R. C. & Storhoff, J. J. A DNA based method for rationally assembling nanoparticles into macroscopic materials. *Nature* **382**, 607–609 (1996).
- Brust, M., Bethell, D., Schiffrin, D. J. & Kiely, C. J. Novel gold-dithiol networks with non-metallic electronic properties. *Adv. Mater.* **7**, 795–797 (1995).
- Murray, C. B., Kagan, C. R. & Bawendi, M. G. Self-organisation of CdSe nanocrystallites into 3-dimensional quantum dot superlattices. *Science* **270**, 1335–1338 (1995).
- Whetten, R. L. *et al.* Nanocrystal gold molecules. *Adv. Mater.* **8**, 428–433 (1996).
- Motte, L., Billoudet, F., Lacaze, E., Douin, J. & Pileni, M. P. Self-organisation into 2D and 3D superlattices of nanosized particles differing by their size. *J. Phys. Chem. B* **101**, 138–144 (1997).
- Pileni, M. P. Nanosized particles made into colloidal assemblies. *Langmuir* **13**, 3266–3276 (1997).
- Brust, M., Walker, M., Bethell, D., Schiffrin, D. J. & Whyman, R. Synthesis of thiol-derivatized gold nanoparticles in a two phase liquid-liquid system. *J. Chem. Soc. Commun.* 801–802 (1994).
- Reetz, M. T., Winter, M. & Tesche, B. Self assembly of tetra-alkylammonium salt-stabilised giant palladium clusters on surfaces. *Chem. Commun.* 147–148 (1997).
- Fink, J., Kiely, C. J., Bethell, D. & Schiffrin, D. J. Self-organisation of nanosized Au particles. *Chem. Mater.* **10**, 922–926 (1998).
- Ohara, P. C., Leff, D. V., Heath, J. R. & Gelbart, W. M. Crystallisation of opals from polydisperse nanoparticles. *Phys. Rev. Lett.* **75**, 3466–3469 (1995).
- Murthy, S., Wang, Z. L. & Whetten, R. L. Thin films of thiol-derivatized gold nanocrystals. *Phil. Mag. Lett.* **75**, 321–327 (1997).
- Luedtke, W. D. & Landman, U. Structure, dynamics and thermodynamics of passivated Au nanocrystallites and their assemblies. *J. Phys. Chem.* **100**, 13323–13329 (1996).
- Terrill, R. H. *et al.* Monolayers in 3D NMR SAXS thermal and electron hopping studies of alkanethiol stabilised Au clusters. *J. Am. Chem. Soc.* **117**, 12537–12548 (1995).
- Badia, A. *et al.* Self-assembled monolayers on gold nanoparticles. *Chem. Eur. J.* **6**, 359–36 (1996).
- Wang, Z. L., Harfenist, S. A., Whetten, R. L., Bentley, J. & Evans, N. D. Bundling and interdigitation of adsorbed thiolate groups in self-assembled nanocrystal lattices. *J. Phys. Chem.* **102**, 3068–3072 (1998).
- Wang, Z. L. Structural analysis of self-assembling nanocrystal superlattices. *Adv. Mater.* **10**, 13–30 (1998).
- Laves, F. *The Theory of Alloy Phases* 124–199 (Am. Society for Metals, Cleveland, Ohio, 1956).
- Sanders, J. V. & Murray, M. J. Ordered arrangements of spheres of two different sizes in opals. *Nature* **275**, 201–202 (1978).
- Sanders, J. V. & Murray, M. J. Close-packed structures of spheres of two different sizes I; Observations on natural opal. *Phil. Mag. A* **42**, 705–720 (1980).
- Yoshimura, S. & Haschisu, S. Order formation in binary mixtures of monodisperse lattices. *Prog. Colloid Polym. Sci.* **68**, 59–70 (1983).
- Haschisu, S. & Yoshimura, S. Optical demonstration of crystalline superstructures in binary mixtures of latex globules. *Nature* **283**, 188–189 (1980).
- Bartlett, P., Ottewill, R. H. & Pusey, P. N. Superlattice formation in binary-mixtures of hard sphere colloids. *Phys. Rev. Lett.* **68**, 3801–3804 (1992).
- Murray, M. J. & Sanders, J. V. Close-packed structures of spheres of two different sizes II; the packing densities of likely arrangements. *Phil. Mag. A* **42**, 721–740 (1980).
- Eldridge, M. D., Madden, P. A. & Frenkel, D. Entropy driven formation of a superlattice in a hard-sphere binary mixture. *Nature* **365**, 35–37 (1993).
- Hume-Rothery, W., Smallman, R. E. & Haworth, C. W. *The Structure of Metals and Alloys* (Metals and Metallurgy Trust, London, 1969).
- Cleveland, C. L., Landman, U., Shafgullin, M. N., Stephens, P. W. & Whetten, R. L. Structural evolution of larger gold clusters. *Z. Phys. D* **40**, 503–508 (1997).
- Hostetler, M. *et al.* Alkanethiolate gold cluster molecules with core diameters from 1.5 to 5.2 nm; monolayer properties as a function of core size. *Langmuir* **14**, 17–30 (1998).
- Herring, C. Effect of change of scale on sintering phenomena. *J. Appl. Phys.* **21**, 301–303 (1950).
- Feldheim, L. & Keating, C. D. Self-assembly of single electron transistors and related devices. *Chem. Soc. Rev.* **27**, 1–12 (1998).

Acknowledgements. We thank Carol Kiely for comments. This work was supported by the EPSRC.

Correspondence and requests for materials should be addressed to C.J.K. (e-mail: kiely@liv.ac.uk).

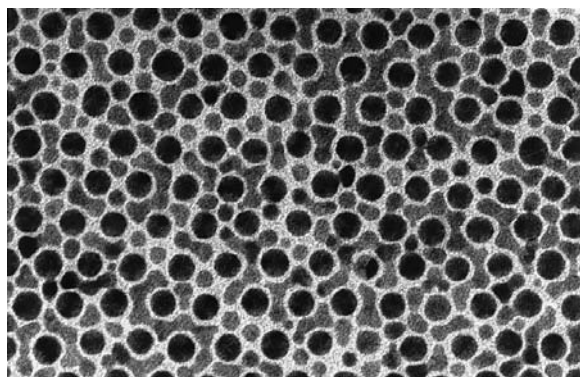


Figure 4 An electron micrograph of a raft of bimodal Au nanoparticles of the type shown in Fig. 1b after storage for two months under ambient conditions. We note that the smaller particles have partially sintered, and formed a semi-continuous network around the larger particles.

Selective amplification by auto- and cross-catalysis in a replicating peptide system

Shao Yao, Indraneel Ghosh, Reena Zutshi & Jean Chmielewski

Department of Chemistry, Purdue University, West Lafayette, Indiana 47907, USA

Self-replication has been demonstrated in synthetic chemical systems based on oligonucleotides^{1–7}, peptides^{8–12} and complementary molecules without natural analogues^{13–16}. However, within a living cell virtually no molecule catalyses its own formation, and the search for chemical systems in which both auto- and cross-catalysis can occur has therefore attracted wide interest¹⁷. One such system, consisting of two self-replicating peptides that catalyse each other's production, has been reported¹⁰. Here we describe a four-component peptide system that is capable of auto- and cross-catalysis and allows for the selective amplification of one or more of the products by changing the reaction conditions. The ability of this system selectively to amplify one or more molecules in response to changes in environmental conditions such as pH or salt concentration supports the suggestion⁸ that self-replicating peptides may have played a role in the origin of life.

We have developed a self-reproducing, autocatalytic as well as cross-catalytic system, in which four different products may be generated from four peptide fragments (Table 1). Selection of one or more of the products would be achieved by altering the reaction conditions within the system. Two of the four potential products, E1E2 and K1K2 were designed as previously described^{11,12,18}. E1E2 forms a coiled-coil only under acidic conditions owing to its large population of glutamate residues¹¹. Moreover, at pH 7.5, E1E2 was designed to form a mixed coiled-coil with K1K2, which alone is also largely a random coil at pH 7.5 (ref. 12). K1K2 contains many lysine residues, and the formation of a coiled-coil can be induced by high concentrations of salts like NaClO₄ at pH 7.5 (ref. 12). The other two peptides, E1K2 and K1E2, were the recombinant hybrids of E1E2 and K1K2, and were designed either to self-associate through antiparallel coiled-coils, or to associate with each other through parallel coiled-coil formation at pH 7.5. These four peptides could be generated from four peptide fragments via the native chemical ligation method developed by Kent¹⁹; two of the fragments, E1 and K1, contain a thioester at their carboxy termini, whereas the other two, E2 and K2, incorporate a free cysteine at their amino termini (Table 1).

Each of the four products may be produced through a different catalytic pathway that would be mediated by the formation of intermediate coiled-coils either with themselves (autocatalysis) or with others (cross-catalysis). Our general design considerations, therefore, focused on these potential auto- and cross-catalytic

pathways (Table 1). For instance, E1E2 and K1K2 may be formed autocatalytically under environmental conditions that suppress repulsion of peptide complexes, such as low pH for E1E2 (ref. 11), and high salt concentrations for K1K2 (ref. 12). The peptides K1E2 and E1K2 could potentially form autocatalytically near neutral pH through the antiparallel templating of their peptide fragments. Cross-catalysis, on the other hand, should occur only if charge complementarity is possible between the template and peptide fragments, such as with K1 and K2 fragments with an E1E2 template (and vice versa), and also with templates containing both positively and negatively charged segments, such as K1E2 and E1K2.

Verification of the overall design was accomplished by running individual reactions containing only two peptide fragments either with or without a templating species (Table 2). As anticipated, reactions with E1E2 and K1K2 as templates proceeded efficiently in both the auto-^{11,12}, and cross-catalytic regimes (Table 2), with cross-catalytic rate constants, k_{cat} , of $3.20 \times 10^4 \text{ M}^{-2} \text{ s}^{-1}$ for K1/K2/E1E2 and $1.60 \times 10^4 \text{ M}^{-2} \text{ s}^{-1}$ for E1/E2/K1K2 at pH 7.5. Cross-catalysis with E1E2 and K1K2 was anticipated because association of these peptides was observed by circular dichroism; experiments demonstrated that a 1:1 mixture of E1E2 and K1K2 was highly helical at pH 7.5, whereas the individual peptides adopted a random coil conformation under the same conditions. The mixed peptide sequences K1E2 and E1K2, however, were only effective templates in their autocatalytic reactions, with autocatalytic rate constants, k_{cat} , of $1.12 \times 10^3 \text{ M}^{-2} \text{ s}^{-1}$ and $8.27 \times 10^3 \text{ M}^{-2} \text{ s}^{-1}$, respectively (Table 2). These data suggested that the K1E2 and E1K2 homo-complexes, which would require antiparallel coiled-coil formation, were favoured. Indeed, circular dichroism spectra of K1E2 and E1K2 showed that as individuals these peptides were highly helical at pH 7.5.

With a clearer understanding of the scope and limitations of auto- and cross-catalysis with these four peptides, the reaction of a mixture of fragments was studied. An equal mixture of E1, E2, K1 and K2 at pH 7.5 produced fairly equivalent amounts of the four potential products, with a preference for E1K2 (Fig. 1a). Addition of each of the four possible peptide templates was next investigated. The templates K1K2 and E1E2 provided the most dramatic shifts in product distribution, going from the initial preference for E1K2 with no added template to preferences for E1E2 (43%) and K1K2 (44%), respectively (Fig. 1a). Addition of K1E2 as a template somewhat increased the relative amounts of K1K2 and K1E2 in the reaction at the expense of E1K2 and E1E2, whereas addition of E1K2 as a template to the reaction mixture led to a modest increase in the production of E1K2 (Fig. 1a).

The addition of individual templating peptides to the mixture of the four peptide fragments did not promote the production of a single product as might have been predicted from our initial design strategy. There were possible reactions between peptides such as E1 and K2 that might be templated by peptides such as K1K2; such reactions had not been anticipated, but might result in reduced selectivities. To determine whether or not these unanticipated catalytic pathways were possible, all additional individual reactions with templates were evaluated. Surprisingly, in many cases template-promoted cross-catalysis was observed (see Table 2 entries in

Table 1 Potential autocatalytic and cross-catalytic reactions of the self-replicating system

	Autocatalysis				Cross-catalysis			
	pH 7.5, 2M NaClO ₄	pH 4.0	pH 7.5		pH 7.5			
Template	K1K2	E1E2	K1E2	E1K2	K1K2	E1E2	K1E2	E1K2
Fragments	K1, K2	E1, E2	K1, E2	E1, K2	E1, E2	K1, K2	E1, K2	K1, E2
	↓	↓	↓	↓	↓	↓	↓	↓
Product	K1K2	E1E2	K1E2	E1K2	E1E2	K1K2	E1K2	K1E2

All peptides were synthesized on the solid support as previously described¹⁹. They contain the following sequences (R = $-(\text{CH}_2)_2\text{CONH}_2$): E1, Ac-ELYALEKELGALKEKLA-COSR; E2, H₂N-CLKELGALEKELYALEK-CONH₂; K1, Ac-KLYALKEKLGALKEKLA-COSR; K2, H₂N-CLKEKLGALKEKLYALKE-CONH₂; K1K2, Ac-KLYALKEKLGALKEKLA-CLKEKLGALKEKLYALKE-CONH₂; K1E2, Ac-KLYALKEKLGALKEKLA-CLKEKLGALKEKELYALEK-CONH₂; E1K2, Ac-ELYALEKELGALKEKLA-CLKEKLGALKEKLYALKE-CONH₂; E1E2, Ac-ELYALEKELGALKEKLA-CLKEKLGALKEKELYALEK-CONH₂.

Table 2 Individual two-peptide coupling reactions

Reaction	Catalytic rate constants k_a ($M^{-2}s^{-1}$) with different templates				Non-catalytic rate constant k_b ($M^{-1}s^{-1}$)
	K1K2	E1E2	K1E2	E1K2	
K1 + K2	NC	3.20×10^4	7.09×10^2	1.18×10^3	1.51×10^{-2}
E1 + E2	1.60×10^4	NC	NC	NC	1.27×10^{-2}
K1 + E2	1.99×10^3	4.18×10^3	1.12×10^3	NC	1.59×10^{-2}
E1 + K2	2.32×10^4	NC	NC	8.27×10^3	4.66×10^{-2}
Dissociation constants (M)					
K1K2	NA	NA	2.38×10^{-8}		
E1E2	1.13×10^{-7}	NA			
K1E2	1.12×10^{-3}	5.57×10^{-5}			
E1K2	1.00×10^{-6}	NA	NA	4.90×10^{-9}	

Above data were obtained by fitting results from individual reactions performed at pH 7.5 to equations described in Methods, using the program SimFit^{4,5}. The non-catalytic rate constant for the reaction between K1 and K2 was the result of three separate simulations (± 0.50). Bold type indicates unanticipated catalytic pathways observed in our system. NC, no catalysis; NA, no association.

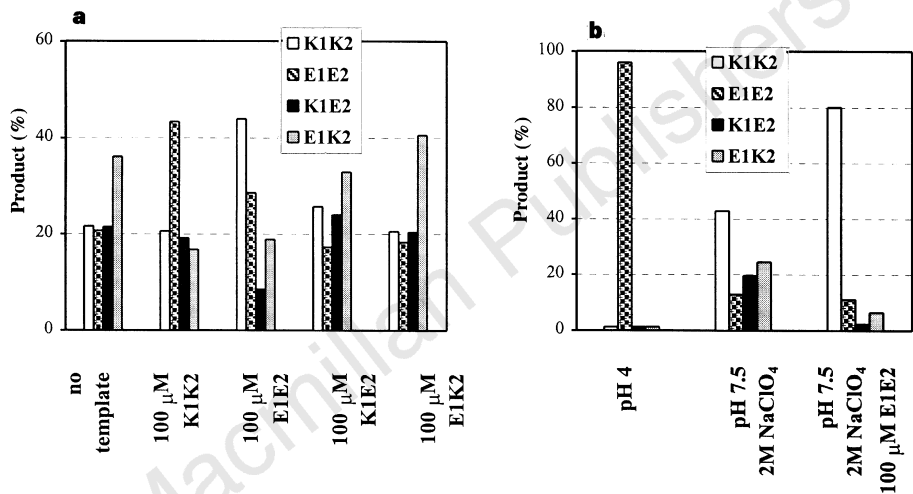


Figure 1 Percentage of each of the four products formed in the mixed fragment reactions ($E1 + E2 + K1 + K2$). **a**, With or without each of the four templates (K1K2, E1E2, K1E2 and E1K2; $100 \mu M$ each) after ~ 500 min at pH 7.5. **b**, Under different reaction conditions: pH 4.0; pH 7.5 in $2M NaClO_4$; and pH 7.5 in $2M NaClO_4$ with $100 \mu M$ E1E2 template.

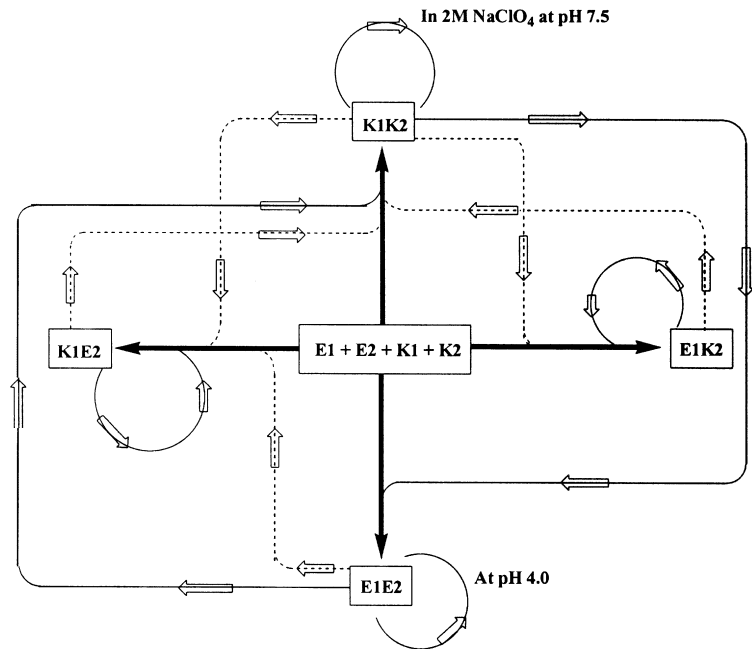


Figure 2 Schematic representation of all reaction pathways uncovered for the mixed fragment reaction $E1 + E2 + K1 + K2$ at pH 7.5. Shown are non-catalytic pathways (bold lines) and catalytic pathways (solid and dashed lines). Solid lines represent pathways anticipated in our design, and dashed lines represent pathways not initially anticipated. Special reaction conditions are noted when used.

bold); K1K2, for instance, catalysed the formation of both K1E2 and E1K2 (and vice versa), whereas E1E2 catalysed the formation of K1E2. The mixed peptides, K1E2 and E1K2, aggregate as trimers as determined by analytical ultracentrifugation, and prefer an anti-parallel orientation as described above. It is possible, therefore, to envisage a dimeric K1E2 template for the production of K1K2 in which the template displays a negatively charged surface for binding the positively charged peptide fragments, leading to a trimeric product complex which contains a 2:1 ratio of K1E2 to K1K2. Similarly, a 1:1 complex between K1E2 and K1K2 could also potentially template K1 and E2, leading to the same trimeric product complex as described above. These data demonstrate that a more complex scheme exists for cross-catalysis than envisaged in the initial design, and may account for the modest selectivities exhibited by the mixed-component reactions.

The information obtained in the individual component reactions was used to analyse the results obtained in the four-component, mixed reactions. The preference for E1K2 in the mixed reaction with no template is reasonable when one considers the following: first, that E1K2 production via non-catalytic and autocatalytic pathways proceeds more rapidly than with the other peptides under these conditions (three and seven times faster, respectively, than the next best peptides), and second, that E1K2 catalyses K1K2 production, which in turn promotes E1K2 production in a cross-catalytic fashion (Table 2). Indeed, simulations of the mixed reaction based on the rate constants and dissociation constants derived from the individual reactions produced a roughly twofold preference for E1K2 in the mixed reaction. Other simulations were also undertaken to evaluate the importance of heterocomplex formation with K1E2 and E1K2; removing the unanticipated catalytic pathways from the calculation (equations (8)–(12) and (16)–(18); see Methods) led to a predicted decrease in the overall percentages of K1E2 and E1K2 produced in the mixture, with an increase in the percentage of E1E2, as compared to the simulation which included these pathways. These simulations point to the significant role of the heterocomplexes between K1E2, E1K2 and the other peptides in the sequence of catalytic events and product distributions in the mixed reactions.

There are also significant trends that can be discerned from the individual rate constants and dissociation constants. For instance, the species with the weakest association (K1K2/K1E2) also has the worst catalytic rate constant (K1/K2/K1E2), whereas the species with the tightest association (E1K2 dimer) does not have the best catalytic rate constant (E1/K2/E1K2). These data make sense if one assumes that the same trends generally apply for the association of the bimolecular (product) and the termolecular (intermediate) complexes; if the bimolecular complex is too stable, product inhibition will lower the apparent rate of catalysis, whereas if the complex is too weak, low levels of the intermediate will lead to reduced catalysis. The best catalysis that is obtained for this particular system occurs for bimolecular species with dissociation constants in the low to sub-micromolar range (K1K2/E1E2 and E1K2/K1K2). It will be of interest, therefore, to see whether this generalization holds true for other replicating peptide systems.

In an effort to enhance the observed selectivities, environmental modifications were employed in the mixed-component reaction (Fig. 1b). Lowering the pH of the reaction mixture to 4.0, for instance, resulted in a highly selective production of E1E2 through its autocatalytic pathway; more than 95% of the product mixture was composed of E1E2. Addition of 2M NaClO₄ to the reaction mixture at pH 7.5, conditions that enhance the autocatalytic production of K1K2 (ref. 12), also significantly altered the product distribution of the reaction mixture with an almost twofold increase in the amount of K1K2 produced, although this peptide still only accounted for less than half of the products (45%). The lower selectivity could be attributed to product inhibition in the autocatalytic pathway, as the dissociation constant for the K1K2 coiled-

coil under these conditions is 0.2 nM (ref. 12). Much higher selectivity was easily accomplished, however, by adding E1E2 to the same reaction mixture, thereby promoting the cross-catalytic pathway for K1K2 production. As a result, more than 80% of the product mixture was composed of K1K2 under these reaction conditions. These results demonstrate that both auto- and cross-catalysis can produce high selectivities when the appropriate conditions are employed.

It has been shown that certain RNA molecules may evolve when challenged with ethidium bromide²⁰. We sought a similar overall strategy in our peptidic system to achieve amplification of one or more molecules within the system through environmental modifications. We have designed, therefore, a complex peptide-based system with the ability to produce four different products from four peptide fragments (Fig. 2). While investigating the scope of this replicating system as a whole, we have demonstrated efficient cross-catalysis in seven unique peptide systems out of the possible 12 cases, and have added two new examples of self-replicating peptides (Table 2). We are able selectively to amplify one or more of the peptides by changing reaction conditions such as pH, and salt or template concentration (Fig. 1b). This finding shows that amplification in a non-enzymatic, self-reproducing system can be induced by environmental changes, thus supporting the suggestion⁸ that peptides and proteins may have been important in the early evolution of living systems. □

Methods

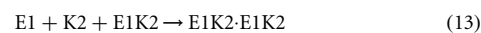
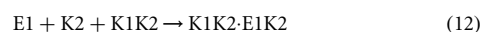
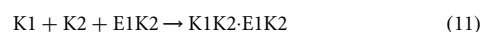
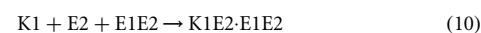
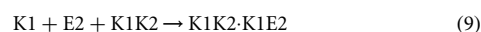
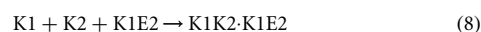
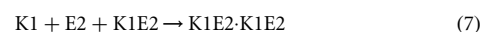
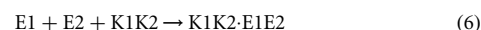
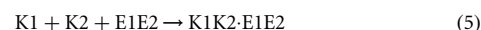
Reaction conditions. Mixtures of peptide fragments (240 μM each, with template when necessary) were incubated at 23 °C in 250 mM MOPS (1% (v/v) 3-mercaptopropionic acid) at the desired pH (4 or 7.5), with or without 2M NaClO₄. At the indicated time interval, 50 μl of the reaction mixture was analysed by HPLC. Reaction products were identified by direct isolation followed by characterization by mass spectrometry, or by co-injection with authentic samples in HPLC. Concentration of products was determined by interpolation of peak area from standard curves constructed from samples with known concentrations.

Reaction models. The following simplified reaction models were used to simulate data obtained from individual reactions at pH 7.5 (refs 4, 5):

Non-catalytic reactions



Catalytic reactions



Dissociation/association





Equations (1), (4)–(6), (14); equations (1), (2), (7)–(10), (15)–(17); and equations (1), (3), (11)–(13), (18), (19) were analysed simultaneously. Results are summarized in Table 2.

Received 8 September 1997; accepted 1 September 1998.

1. von Kiedrowski, G. A self-replicating hexadeoxynucleotide. *Angew. Chem. Int. Edn Engl.* **25**, 932–935 (1986).
2. von Kiedrowski, G., Wlotzka, B. & Helbing, J. Sequence dependence of template-directed syntheses of hexadeoxynucleotide derivatives with 3'-5' pyrophosphate linkage. *Angew. Chem. Int. Edn Engl.* **28**, 1235–1237 (1989).
3. von Kiedrowski, G., Wlotzka, B., Helbing, J., Matzen, M. & Jordan, S. Parabolic growth of a self-replicating hexadeoxynucleotide bearing a 3'-5' phosphoamidate linkage. *Angew. Chem. Int. Edn Engl.* **30**, 423–426 (1991).
4. Sievers, D. & von Kiedrowski, G. Self-replication of complementary nucleotide-based oligomers. *Nature* **369**, 221–224 (1994).
5. Achilles, T. & von Kiedrowski, G. A self-replicating system from three starting materials. *Angew. Chem. Int. Edn Engl.* **32**, 1198–1201 (1993).
6. Zielinski, W. S. & Orgel, L. E. Autocatalytic synthesis of a tetranucleotide analogue. *Nature* **327**, 346–347 (1987).
7. Li, T. & Nicolaou, K. C. Chemical self-replication of palindromic duplex DNA. *Nature* **369**, 218–221 (1994).
8. Lee, D. H., Granja, J. R., Martinez, J. A., Severin, K. & Ghadiri, M. R. A self-replicating peptide. *Nature* **382**, 525–528 (1996).
9. Severin, K., Lee, D. H., Kennan, A. J. & Ghadiri, M. R. A synthetic peptide ligase. *Nature* **389**, 706–709 (1997).
10. Lee, D. H., Granja, J. A., Severin, K., Yokobayashi, Y. & Ghadiri, M. R. Emergence of symbiosis in peptide self-replication through a hypercyclic network. *Nature* **390**, 591–594 (1997).
11. Yao, S., Ghosh, I., Zutshi, R. & Chmielewski, J. A pH-modulated, self-replicating peptide. *J. Am. Chem. Soc.* **119**, 10559–10560 (1997).
12. Yao, S., Ghosh, I., Zutshi, R. & Chmielewski, J. A self-replicating peptide under ionic control. *Angew. Chem. Int. Edn Engl.* **37**, 478–481 (1998).
13. Tjivikua, T., Ballester, P. & Rebek, J. Jr A self-replicating system. *J. Am. Chem. Soc.* **112**, 1249–1250 (1990).
14. Nowick, J. S., Feng, Q., Tjivikua, T., Ballester, P. & Rebek, J. Jr Kinetic studies and modeling of a self-replicating system. *J. Am. Chem. Soc.* **113**, 8831–8839 (1991).
15. Wintner, E. A., Conn, M. M. & Rebek, J. Jr Self-replicating molecules: a second generation. *J. Am. Chem. Soc.* **116**, 8877–8884 (1994).
16. Wintner, E. A., Tsao, B. & Rebek, J. Jr Evidence against an alternative mechanism for a self-replicating system. *J. Org. Chem.* **60**, 7997–8001 (1995).
17. Orgel, L. E. Unnatural selection in chemical systems. *Acc. Chem. Res.* **28**, 109–118 (1995).
18. Zhou, N. E., Kay, C. M. & Hodges, R. S. The role of interhelical ionic interactions in controlling protein folding and stability. *De novo* designed synthetic two-stranded alpha-helical coiled-coils. *J. Mol. Biol.* **237**, 500–512 (1994).
19. Dawson, P. E., Muir, T. W., Clark-Lewis, I. & Kent, S. B. Synthesis of proteins by native chemical ligation. *Science* **266**, 776–779 (1994).
20. Kramer, F. R., Mills, D. R., Cole, P. E., Nishihana, T. & Spiegelman, S. Evolution in vitro: sequence and phenotype of a mutant RNA resistant to ethidium bromide. *J. Mol. Biol.* **89**, 719–736 (1974).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. This work was supported by the NSF.

Correspondence and requests for materials should be addressed to J.C. (e-mail: chml@cv1.chem.purdue.edu).

A new model for Proterozoic ocean chemistry

D. E. Canfield

Biological Institute, Odense University, Campusvej 55, 5230 Odense M, Denmark

There was a significant oxidation of the Earth's surface around 2 billion years ago (2 Gyr)^{1–4}. Direct evidence for this oxidation comes, mostly, from geological records of the redox-sensitive elements Fe and U reflecting the conditions prevailing during weathering^{1–3}. The oxidation event was probably driven by an increased input of oxygen to the atmosphere arising from an increased sedimentary burial of organic matter between 2.3 and 2.0 Gyr⁵. This episode was postdated by the final large precipitation of banded iron formations around 1.8 Gyr^{1,2}. It is generally believed that banded iron formations precipitated from an ocean

whose bottom waters contained significant concentrations of dissolved ferrous iron, and that this sedimentation process terminated when aerobic bottom waters developed, oxidizing the iron and thus removing it from solution^{1,2}. In contrast, I argue here that anoxic bottom waters probably persisted until well after the deposition of banded iron formations ceased; I also propose that sulphide, rather than oxygen, was responsible for removing iron from deep ocean water. The sulphur-isotope record supports this hypothesis as it indicates increasing concentrations of oceanic sulphate, starting around 2.3 Gyr⁶, leading to increasing rates of sulphide production by sulphate reduction. The increase in sulphide production became sufficient, around 1.8 Gyr, to precipitate the total flux of iron into the oceans. I suggest that aerobic deep-ocean waters did not develop until the Neoproterozoic era (1.0 to ~0.54 Gyr), in association with a second large oxidation of the Earth's surface. This new model is consistent with the emerging view of Precambrian sulphur geochemistry and the chemical events leading to the evolution of animals, and it is fully testable by detailed geochemical analyses of preserved deep-water marine sediments.

The oxygen content of deep ocean waters is controlled mainly by the balance between the supply of oxygen from downwelling surface waters and oxygen consumption by decomposing organic debris⁷. The oxidation of submarine hydrothermal inputs, by comparison, requires 2 to 3 orders of magnitude less oxygen than the oxidation of organic matter⁸. In the modern ocean, the downwelling of cold high-latitude surface waters saturated with oxygen results in largely aerobic bottom waters⁷. Furthermore, primary production, fuelled by nutrients dissolved in upwelling deep ocean water, produces insufficient organic matter to exhaust the oxygen in deep waters as the organic matter sinks and decomposes⁷. Deep-water anoxia could occur, however, by a reduction in the concentration of oxygen in downwelling surface waters, an increase in the nutrient concentration of the oceans (allowing more primary production, and consequently a greater organic carbon rain to the deep ocean, at equivalent rates of ocean ventilation), or by an increase in the efficiency of nutrient utilization, which is particularly poor in the modern high-latitude ocean⁷.

A simple three-box model, used primarily with respect to the ice-age ocean, has been useful in exploring the regulation of the oxygen content of deep ocean waters⁷. The ocean is divided into a deep ocean box (d), and two surface boxes: one representing the region of bottom water formation (f) and the other the remainder of the surface ocean (s). In simplified form, the concentration of oxygen in the deep ocean is given by⁷:

$$\text{O}_2(\text{d}) = \text{O}_2(\text{f}) - r[\text{P}(\text{d}) - \text{P}(\text{f})] \quad (1)$$

where O_2 is oxygen concentration (micromolar, μM), P is phosphate concentration (μM), with phosphate taken as the limiting nutrient in primary production, and r is the Redfield ratio between O_2 consumption and phosphate production during organic matter mineralization, taken here as 169 (ref. 7). Estimates for r vary between 120 and 200 (ref. 9); in this range, however, the value of this parameter does not alter my major conclusions. The difference $[\text{P}(\text{d}) - \text{P}(\text{f})]$ represents the amount of phosphorus in upwelling deep water that is used during primary production.

The model can be used to explore how the oxygen content of deep ocean waters ($\text{O}_2(\text{d})$) responds to variable levels of both surface water oxygen (represented by $\text{O}_2(\text{f})$) and nutrient availability $[\text{P}(\text{d}) - \text{P}(\text{f})]$ (Fig. 1). A drop in $\text{O}_2(\text{f})$, for example, means that a smaller value of $[\text{P}(\text{d}) - \text{P}(\text{f})]$ is required to maintain aerobic conditions in the deep ocean. In physical terms, lower nutrient availability reduces primary production, and therefore lowers oxygen demand in the deep ocean by decomposing sunken organic matter. This is why lower nutrient availability allows aerobic conditions to persist as values of $\text{O}_2(\text{f})$ drop. When $\text{O}_2(\text{f})$ falls to 25 μM , aerobic conditions are nearly impossible to maintain when there is any appreciable phosphorus in the oceans. With the present-

day $[P(d) - P(f)]$ value of $0.91 \mu\text{M}$ (ref. 7), concentrations of $\text{O}_2(f)$ exceeding $150\text{--}200 \mu\text{M}$ are required to maintain even minimally aerobic deep ocean conditions. The concentration of $\text{O}_2(f)$ in the modern ocean is $\sim 325 \mu\text{M}$ (ref. 7), which is in approximate equilibrium with the atmosphere. Water-column primary production and air injection can drive $\text{O}_2(f)$ to higher than equilibrium levels, although oversaturation rarely exceeds $10\text{--}20 \mu\text{M}$ O_2 (ref. 10), and may be much less.

This analysis shows that the development of aerobic deep ocean waters requires a combination of either low atmospheric oxygen levels and extremely low nutrient availability, or high atmospheric oxygen levels and nutrient availability comparable to the present-day ocean. The exogenic inventory of phosphorus has probably not varied greatly over much of geological time, meaning that levels of nutrient availability have also not varied greatly². Therefore, Proterozoic ($2.5\text{--}0.54$ Gyr) oceans probably had phosphorus levels similar to the modern ocean. Under these circumstances, aerobic deep ocean water would have required high levels of atmospheric oxygen—within a factor of 2 or 3 of the present level (Fig. 1).

Several lines of evidence, however, suggest that atmospheric oxygen did not reach these high levels until the Neoproterozoic era, considerably later than the last period of major banded iron formation (BIF) deposition. The evidence includes the late Proterozoic rise of animals^{1,11} and colourless sulphur bacteria¹², and the large increase in stable sulphur isotope fractionation at $0.6\text{--}1.0$ Gyr¹². All of these have been linked to oxygen levels rising to $>10\text{--}20\%$ of present-day levels^{1,11–13}, and are roughly coincident with a second major input pulse of oxygen to the atmosphere¹⁴.

Here I propose an alternative model for the development of ocean chemistry during the Precambrian aeon, which is consistent with a two-stage rise in atmospheric oxygen levels^{1,2,4,11}. The first stage, at $\sim 2.3\text{--}2.0$ Gyr, produced enough oxygen to influence the chemical behaviour of many redox-sensitive elements during weathering, but not enough to create an aerobic deep ocean. During this period,

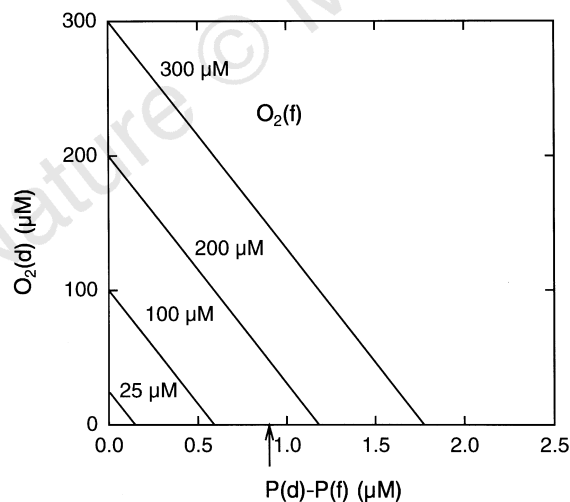


Figure 1 The relationship between deep-water oxygen concentration ($\text{O}_2(d)$) and phosphorus availability (represented by $[P(d) - P(f)]$) in the oceans. Solid lines show constant surface-water oxygen levels ($\text{O}_2(f)$). The model results have been calculated from equation (1), assuming a three-box ocean and well-mixed boxes⁷. The figure shows, for example, that with $300 \mu\text{M}$ $\text{O}_2(f)$, deep-water anoxia occurs when phosphorus availability rises to $\sim 1.6 \mu\text{M}$. With no nutrients in the ocean, bottom water oxygen remains the same as surface values. Present-day phosphorus availability is $\sim 0.9 \mu\text{M}$ (arrow); this requires $\text{O}_2(f)$ values of $\geq 160\text{--}170 \mu\text{M}$ to maintain oxygen in deep-ocean waters. The three-box model is a highly simplified picture of ocean dynamics which only gives broad averages. The absence of information on Precambrian ocean circulation, however, justifies this simplified approach to understanding the relationship between atmospheric oxygen levels and Precambrian ocean anoxia. See text for details.

seawater sulphate levels increased, stimulating sulphate reduction and generating enough sulphide to remove dissolved ferrous iron as iron sulphides. The second stage, in the Neoproterozoic era, generated enough oxygen to sweep the deep ocean of sulphide, and resulted in the many biological and geochemical innovations that occurred at this time.

A transition from an Fe-rich to a sulphidic early Proterozoic ocean is consistent with the isotope record of non-hydrothermal sedimentary sulphides (Fig. 2; see also ref. 6). A prominent feature of this record is a large increase in the isotope difference between seawater sulphate and sulphide at $\sim 2.20\text{--}2.30$ Gyr^{6,15}. The small fractionations before $2.20\text{--}2.30$ Gyr are generally comparable to the small ($\leq 4\%$) fractionations reported for bacterial sulphate reduction with ≤ 1 mM sulphate¹⁶. Therefore, 1 mM may be a reasonable upper limit for the concentration of seawater sulphate before $2.20\text{--}2.30$ Gyr^{6,15} (for a different view, see ref. 17). The increase in fractionation at $2.20\text{--}2.30$ Gyr is thus consistent with a rise in seawater sulphate concentration to >1 mM, where much larger fractionations ($4\text{--}46\%$) by sulphate-reducing bacteria are expressed^{15,18}. An increase in sulphate concentration at $2.20\text{--}2.30$ Gyr would also coincide approximately with the timing of the first stage of Earth surface oxidation³. It is possible that in some restricted ocean basins before $2.2\text{--}2.3$ Gyr—where, for example, mineral sulphate deposition occurred, or anoxygenic photosynthesis was particularly active—there might have been elevated concentrations of sulphate¹⁵. Under these circumstances, we might expect there to have been enhanced isotope fractionation during sulphate reduction, although the isotope record provides few, if any, compelling examples (Fig. 2).

Models and empirical observations of sedimentary sulphur contents both show that low concentrations of sulphate, for example the

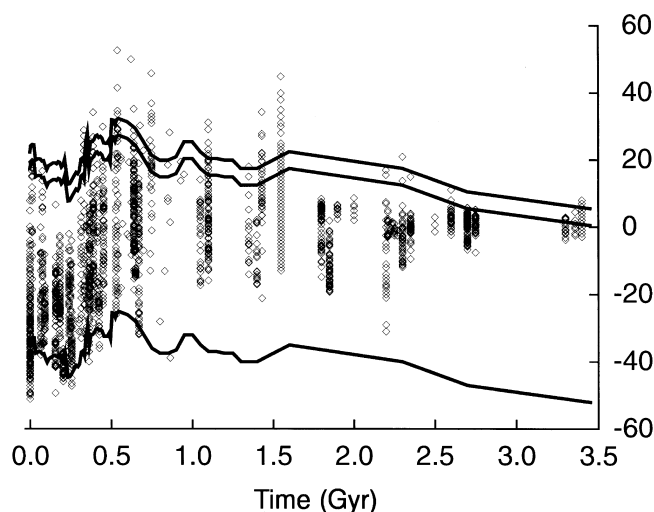


Figure 2 The isotope composition of sedimentary sulphides of probable biological origin over geological time. This excludes sulphides of obviously hydrothermal origin, those hosted in non-sedimentary rocks, and those from sediments that have experienced beyond greenschist phase metamorphism. Shown for comparison is the isotope composition of seawater sulphate, mostly obtained from measuring the isotope composition of evaporitic anhydrite or gypsum. Before 2.2 Gyr the isotope composition of sulphate is determined mostly from barites. The two upper lines show the isotope composition of sulphate. The range (5%) represented by the pair of lines shows the uncertainty in the inferred isotopic composition of seawater sulphate. However, before 1 Gyr, knowledge of the isotope composition of seawater sulphate is constrained by a lack of data, resulting in a loss of detail. The lower line shows the isotope composition of sulphate displaced by 55‰ to more ^{34}S -depleted values. Comparison of the isotope composition of sulphides and this lower line demonstrates how the fractionation between sulphate and sedimentary sulphides has increased over time. Data are taken partly from an earlier compilation¹²; additional data sources are available as Supplementary Information.

50–500 μM typically found in freshwater lakes, limits the rate of sulphate reduction^{19,20}. An increase in the sulphate concentration of low sulphate systems will therefore result in increased rates of sulphate reduction. Once the rate of sulphate reduction exceeded the rate at which sulphide-reactive Fe was delivered to early Proterozoic oceans, dissolved Fe would have been completely removed from the oceans. It is possible that sulphate concentrations rose gradually between 2.3 Gyr, when high rates of carbon burial first commenced⁴, and 1.8 Gyr, when the last of the major BIFs were deposited². This would mean that levels of sulphate first exceeded those required to produce large isotope fractionations at 2.2–2.3 Gyr, whereas sulphate reduction rates first exceeded rates of Fe delivery to the oceans at ~ 1.8 Gyr.

In the Neoproterozoic era, there was a return to a much smaller-scale deposition of BIFs²¹, for example those found in the Rapitan Group (~ 0.70 – 0.75 Gyr) of the Mackenzie Mountains, Canada. These BIFs are intimately associated with glacial deposits, and they were probably deposited during periods of ocean stagnation promoted by glaciation²¹. It is not known whether the dissolved Fe that precipitated as BIFs accumulated throughout the ocean basin, or whether it was only present in regions analogous to modern-day oxygen minimum zones. In either case, the accumulation of dissolved Fe rather than sulphide shows that there were low rates of sulphate reduction, compared to rates of Fe delivery, within broad anoxic regions of the ocean. A possible explanation is a reduction of sulphate concentration during ocean stagnation, to levels inhibiting sulphate reduction, similar to the conditions found in the early Proterozoic. A low water-column concentration of sulphate during the Neoproterozoic era has previously been advocated²² to explain the preponderance of ^{34}S -enriched sedimentary sulphides found from this time. It is possible that the concentration of seawater sulphate did not rise to near present levels until after the second stage of Earth-surface oxidation in the Late Neoproterozoic era.

In the modern oceans, there is enough sulphide produced by sulphate reduction to precipitate about seven times the delivery flux of iron to the oceans (Table 1). This does not occur, and oxidized iron persists in many modern marine sediments, because there is intense sulphide oxidation (by oxygen or other oxidized compounds) in sediments¹². However, with modern-day fluxes of reactive Fe and rates of sulphate reduction, dramatically reduced atmospheric oxygen would promote ocean anoxia and the development of sulphidic ocean conditions. Similar conditions of excess sulphide production with respect to Fe delivery are envisaged for middle to late Proterozoic oceans. Late Proterozoic glacial intervals are an exception, where ocean stagnation is proposed to have drawn down sulphate levels so that rates of sulphate reduction were less than those of Fe delivery, leading to renewed BIF deposition.

Table 1 Comparison of delivery flux of reactive iron and sulphate reduction rate in the modern ocean

Reactive Fe from terrigenous sources	$4.9 \times 10^{12} \text{ mol yr}^{-1}$
Hydrothermal Fe input	$1.7 \times 10^{11} \text{ mol yr}^{-1}$
Total reactive Fe delivery flux	$5.1 \times 10^{12} \text{ mol yr}^{-1}$
Maximum sulphide removal as FeS_2	$10.2 \times 10^{12} \text{ mol yr}^{-1}$
Sulphate-reduction rate	$73 \times 10^{12} \text{ mol yr}^{-1}$

I computed the delivery flux of sulphide-reactive Fe to the oceans by combining the average concentration of total Fe in river particulates (4.4 wt%)²³, and the proportion of this total Fe that is sulphide-reactive (0.27; ref. 25), with the flux of terrigenous particulates to ocean. The composition of river particulates is assumed to represent terrigenous material in general. The total flux of terrigenous particulates ($23 \times 10^{15} \text{ g yr}^{-1}$) consists of a river-derived component³⁰ of $20 \times 10^{15} \text{ g yr}^{-1}$ and $1.5 \times 10^{15} \text{ g yr}^{-1}$ each from glacial discharge and wind-blown dust. The terrigenous particle flux is added the input from mid-ocean ridge hydrothermal vents, computed by combining the average concentration of Fe in vent fluids with the fluid flow through the high-temperature vents⁸. The potential maximum removal rate of sulphide (as pyrite, FeS_2) is equivalent to twice the delivery flux of Fe to the oceans. Global rates of sulphate reduction are calculated here from a literature compilation of 220 sulphate-reduction rate determinations. Rates were organized into eight different sedimentary environments, including shallow intertidal, coastal upwelling, estuaries and bays, high deposition shelf, low deposition shelf, slope, rise and abyss. Rates were averaged and then multiplied by the ocean area for each environment to yield total rates of sulphate reduction.

I believe that there should be direct evidence for a middle- to late-Proterozoic sulphide-rich ocean in the geochemical record of preserved deep-water marine sediments. Unfortunately, there are very few geochemical reports from this time that can be used to explore for euxinic conditions. A non-clastic source of Fe has been identified in the deep-water black shales of the Newland Formation of the Mesoproterozoic (1.35–1.4 Gyr) Belt Supergroup in the United States²³. A similar non-clastic Fe source has been identified in the laminated euxinic sediments of the Black Sea²⁴, and this non-clastic Fe, associated with sulphide, requires a euxinic basin for its deposition^{24,25}. Thus, the shales of the Newland Formation could have been deposited under euxinic conditions, but further analysis is required to substantiate this conclusion. Euxinic conditions have also been proposed during the deposition of the Mesoproterozoic Nonesuch Shale (1.1 Gyr) in the United States²⁶, for some intervals of the Velkerri Formation (1.45 Gyr), McArthur Basin, Australia²⁷, and for black shales from the Upper Proterozoic Hedmark Group, Norway²⁸.

These few reports should not, however, be taken as proof of a sulphide-rich ocean in the middle to late Proterozoic; my study should be considered as a basis for investigating middle Proterozoic sediment geochemistry, with a view to understanding depositional conditions. However, the development of a middle-Proterozoic sulphidic ocean, as proposed here, best explains the available geochemical and biological information on the course of Earth-surface change during the Proterozoic aeon. □

Received 23 October 1997; accepted 17 September 1998.

- Cloud, P. E. Jr A working model of the primitive Earth. *Am. J. Sci.* **272**, 537–548 (1972).
- Holland, H. D. *The Chemical Evolution of the Atmosphere and Oceans* (Princeton Univ. Press, Princeton, 1984).
- Holland, H. D. & Beukes, N. A paleoweathering profile from Griqualand West, South Africa: evidence for a dramatic rise in atmospheric oxygen between 2.2 and 1.9 bybp. *Am. J. Sci.* **290-A**, 1–34 (1990).
- Des Marais, D. J., Strauss, H., Summons, R. E. & Hayes, J. M. Carbon isotope evidence for the stepwise oxidation of the Proterozoic environment. *Nature* **359**, 605–609 (1992).
- Karhu, J. A. & Holland, H. D. Carbon isotopes and the rise of atmospheric oxygen. *Geology* **24**, 867–870 (1996).
- Hayes, J. M., Lambert, I. B. & Strauss, H. in *The Proterozoic Biosphere: A Multidisciplinary Study* (eds Schopf, J. W. & Klein, C.) 129–134 (Cambridge Univ. Press, 1992).
- Sarmiento, J. L., Herbert, T. D. & Toggweiler, J. R. Causes of anoxia in the world ocean. *Glob. Biogeochem. Cycles* **2**, 115–128 (1988).
- Ridge/vents Workshop Group. In *Global Impact of Submarine Hydrothermal Processes* (eds Kadko, D., Baker, E., Alt, J. & Baross, J.) 4–15 (NSF RIDGE Initiative and NOAA Vents Program, 1994).
- Shaffer, G. Biogeochemical cycling in the global ocean 2. New production, Redfield ratios, and remineralization in the organic pump. *J. Geophys. Res.* **101**, 3723–3745 (1996).
- Emerson, S., Quay, P. D., Stump, C. & Schudlich, R. Chemical tracers of productivity and respiration in the subtropical Pacific Ocean. *J. Geophys. Res.* **100**, 15873–15887 (1995).
- Knoll, A. H. in *Origin and Early Evolution of the Metazoa* (eds Lipps, J. H. & Signor, P. W.) 53–84 (Plenum, New York, 1992).
- Canfield, D. E. & Teske, A. Late proterozoic rise in atmospheric oxygen concentration inferred from phylogenetic and sulphur-isotope studies. *Nature* **382**, 127–132 (1996).
- Berkner, L. V. & Marshall, L. C. On the origin and rise of oxygen concentration in the Earth's atmosphere. *J. Atmos. Res.* **22**, 225–261 (1965).
- Knoll, A. H., Hayes, J. M., Kaufman, A. J., Swett, K. & Lambert, I. B. Secular variation in carbon isotope ratios from Upper Proterozoic successions of Svalbard and East Greenland. *Nature* **321**, 832–838 (1986).
- Cameron, E. M. Sulphate and sulphate reduction in early Precambrian oceans. *Nature* **296**, 145–148 (1982).
- Harrison, A. G. & Thode, H. G. Mechanisms of the bacterial reduction of sulfate from isotope fractionation studies. *Trans. Faraday Soc.* **53**, 84–92 (1958).
- Ohmoto, H., Kakegawa, T. & Lowe, D. R. 3.4-billion-year-old biogenic pyrites from Barberton, South Africa: Sulfur isotope evidence. *Science* **262**, 555–557 (1993).
- Habicht, K. S. & Canfield, D. E. Sulphur isotope fractionation in modern microbial mats and the evolution of the sulphur cycle. *Nature* **382**, 342–343 (1996).
- van Cappellen, P. & Wang, Y. in *Metal Contaminated Sediments* (ed. Allen, H. E.) 21–64 (Ann Arbor, Chelsea, Michigan, 1995).
- Berner, R. A. & Raiswell, R. Burial or organic carbon and pyrite sulfur in sediments over geologic time. *Geochim. Cosmochim. Acta* **47**, 855–862 (1983).
- Beukes, N. J. & Klein, C. in *The Proterozoic Biosphere: A Multidisciplinary Study* (eds Schopf, J. W. & Klein, C.) 147–151 (Cambridge Univ. Press, 1992).
- Logan, G. A., Hayes, J. M., Hieshima, G. B. & Summons, R. E. Terminal Proterozoic re-organization of biogeochemical cycles. *Nature* **376**, 53–56 (1995).
- Schieber, J. Anomalous iron distribution in shales as a manifestation of “non-clastic iron” supply to sedimentary basins: relevance for pyritic shales, base-metal mineralization, and oolitic ironstone deposits. *Mineral. Deposita* **30**, 294–302 (1995).
- Canfield, D. E., Lyons, T. W. & Raiswell, R. A model for iron deposition to euxinic Black Sea sediments. *Am. J. Sci.* **296**, 818–834 (1996).
- Raiswell, R. & Canfield, D. E. Sources of iron for pyrite formation in marine sediments. *Am. J. Sci.* **298**, 219–245 (1998).
- Imbus, S. W., Macko, S., Elmore, R. D. & Engel, M. Stable isotope ($\text{C}_2\text{S}_2\text{N}$) and molecular studies on the Precambrian Nonesuch Shale (Wisconsin–Michigan, USA): evidence for differential preservation rates, depositional environment and hydrothermal influence. *Chem. Geol.* **101**, 255–281 (1992).
- Jackson, M. J. & Raiswell, R. Sedimentology and carbon–sulphur geochemistry of the Velkerri Formation, a mid-Proterozoic potential oil source in northern Australia. *Precamb. Res.* **54**, 81–108 (1991).

28. Vidal, G. & Nystuen, J. P. Micropaleontology, depositional environment, and biostratigraphy of the Upper Proterozoic Hedmark Group, Southern Norway. *Am. J. Sci.* **290-A**, 170–211 (1990).
29. Canfield, D. E. The geochemistry of river particulates from the continental United States: major elements. *Geochim. Cosmochim. Acta* **61**, 3349–3365 (1997).
30. Milliman, J. D. & Syvitski, J. P. M. Geomorphic/tectonic control of sediment discharge to the ocean: the importance of mountainous rivers. *J. Geol.* **100**, 525–544 (1992).

Supplementary information is available on Nature's World-Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. I thank J. Hayes, B. Thamdrup, D. Des Marais and A. Knoll for comments on the manuscript. Financial support came from the Danish National Science Research Council (SNF) and the Danish National Research Foundation (Danmarks Grundforskningsfond). This Letter is dedicated to the memory of H. Jannasch.

Correspondence and requests for materials should be addressed to D.E.C. (e-mail: dec@biology.ou.dk).

Low-latitude glaciation and rapid changes in the Earth's obliquity explained by obliquity–oblateness feedback

Darren M. Williams*, James F. Kasting† & Lawrence A. Frakes‡

* School of Science, Penn State Erie, The Behrend College, Station Road, Erie, Pennsylvania 16563-0203, USA

† Department of Geosciences, The Pennsylvania State University, 443 Deike Building, University Park, Pennsylvania 16802, USA

‡ Department of Geology and Geophysics, University of Adelaide, Adelaide, South Australia 5005, Australia

Palaeomagnetic data suggest that the Earth was glaciated at low latitudes during the Palaeoproterozoic^{1,2} (about 2.4–2.2 Gyr ago) and Neoproterozoic^{3–8} (about 820–550 Myr ago) eras, although some of the Neoproterozoic data are disputed^{9,10}. If the Earth's magnetic field was aligned more or less with its spin axis, as it is today, then either the polar ice caps must have extended well down into the tropics—the 'snowball Earth' hypothesis⁸—or the present zonation of climate with respect to latitude must have been reversed. Williams¹¹ has suggested that the Earth's obliquity may have been greater than 54° during most of its history, which would have made the Equator the coldest part of the planet¹². But this would require a mechanism to bring the obliquity down to its present value of 23.5°. Here we propose that obliquity–oblateness feedback¹³ could have reduced the Earth's obliquity by tens of degrees in less than 100 Myr if the continents were situated so as to promote the formation of large polar ice sheets. A high obliquity for the early Earth may also provide a natural explanation for the present inclination of the lunar orbit with respect to the ecliptic (5°), which is otherwise difficult to explain.

The most straightforward explanation for low-latitude glaciation is simply that the climate was very cold in Precambrian times as a consequence of reduced solar luminosity¹⁴, uncompensated by high levels of greenhouse gases. It is difficult to explain, however, how the ice caps could have extended to low latitudes without causing the extinction of most or all surface life. Energy-balance climate models^{14,15} suggest that runaway glaciation should occur if the ice line extends equatorwards of 25–30° latitude. At least some of the ancient glaciations appear to have occurred at latitudes much lower than this⁸. Had the ice line moved down below the critical latitude for stability, the Earth should have switched into a globally glaciated state in which even the oceans would have frozen down to a depth of ~1 km, provided that the thickness was limited by the efficiency of heat flow through the ice. If this had occurred during the Neoproterozoic era, the Earth could eventually have escaped from this state

by atmospheric build-up of volcanically released CO₂, which would have warmed the surface and melted the ice¹⁵. However, this process would have required ~10⁶–10⁷ years, during which time the ocean and all marine organisms would have been deprived of sunlight. This is in apparent conflict with evidence for the continuous presence of photosynthetic marine organisms in the geological record, although some authors⁸ believe that this is exactly what happened.

The high-obliquity hypothesis¹¹ of Williams can resolve this climate paradox because high latitudes could have remained ice-free while the tropics were glaciated. A potentially damaging counter-argument has been presented by Vanyo and Awramik¹⁶, based on the sinusoidal growth pattern of an 850-Myr-old stromatolite which suggests a planetary obliquity of 26.5° at that time. This argument presumes that the stromatolite growth face remained perpendicular to the Sun's noontime rays throughout the course of the year. Heliotropism has indeed been seen in modern columnar stromatolites in Australia and Yellowstone¹⁷, but in no case has it been shown that the growth pattern accurately tracks the minimum daily solar zenith angle over the year. Hence the main argument for a low obliquity during the late Proterozoic seems inconclusive.

For the high-obliquity hypothesis to be viable, one needs to identify a mechanism that could have caused the Earth's obliquity to decrease by several tens of degrees between ~600 Myr ago—the age of the youngest low-latitude glacial deposits^{3–5}—and 430 Myr ago, when palaeotidal data suggest that the obliquity was close to its present value¹¹. Williams himself suggested that core–mantle dis-

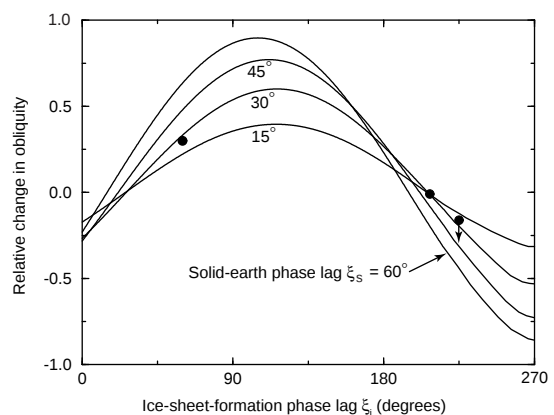


Figure 1 Relative rate of obliquity drift as a function of a ice-sheet-formation phase lag. Averaged over a precession cycle, the rate of obliquity drift may be written^{13,22} $d\theta/dt = K(\Delta_2/J_2)[\sin(\xi_i) - f(\xi_s + \xi_i)]$, where θ is the mean obliquity, and the constant K is a function of various dynamical and geophysical parameters appearing in the equations of precession^{12,22}. The ellipticity coefficient J_2 is equal to $(C - A)/M_\oplus R_\oplus^2$, where C and A are Earth's principal inertial moments. The planetary oblateness is assumed to vary sinusoidally and to lag the phase of the obliquity–insolation cycle by an angle ξ_s . (In reality, ice-sheet response to insolation may be much more complicated.) The amplitude of the oblateness variation, Δ_2/J_2 , represents the maximum change that could ever be realized by lowering sea level and loading Earth's continents with ice. A second oblateness variation, which opposes the first, results from isostatic adjustment of the solid Earth and is assumed to lag ice volume by ξ_s . The long response time (~10⁴ yr) of the viscous upper mantle to changes in ice volume allows the solid-Earth variation to cancel only a fraction, $f(\xi_s)\Delta_2/J_2$, of the water-ice variation, where the fractional amplitude may be written²² $f(\xi_s) = 0.8 - 0.8\xi_s/90^\circ$. Shown are the rate and direction of obliquity drift for different values of the phase lags ξ_i and ξ_s . Plotted on the vertical axis is the quantity $[\sin(\xi_i) - f(\xi_s)\sin(\xi_s + \xi_i)]$. The filled circles mark the values of ξ_i used for the three calculations shown in Fig. 3. The solid-Earth phase lag ξ_s , which depends in the period of the obliquity oscillation in the Precambrian times (ref. 22), was initialized to 28° for each run. For the case when $\xi_i = 230^\circ$, the value ξ_s grows from 28° to 45° over 100 Myr (indicated by an arrow) as a result of an increase to the ice-loading frequency as the obliquity drifts downward.

sipation could have caused the obliquity to decrease. However, this is unlikely if the viscosity of the outer core is low, as many have suggested (compare refs 11, 18). Furthermore, even if it could be shown to work, this mechanism should have operated throughout the Earth's history, making it difficult to explain how the obliquity could have remained high as late as 600 Myr ago.

An obliquity-reducing mechanism that might have operated preferentially during the late Proterozoic is obliquity–oblateness feedback¹³, sometimes termed ‘climate friction’. A secular obliquity drift can occur as a result of the time delay between a planet's obliquity oscillation, which results from precession¹², and obliquity–(insolation–) driven variations in planetary oblateness caused by changes to continental ice volume and sea level (see Fig. 1). The oblateness itself affects the rate at which the obliquity varies with time, thereby completing the obliquity–oblateness feedback loop.

Rubincam¹³ demonstrated that such climate friction could account for a +10° drift in the obliquity of Mars over 4.5 Gyr. Bills¹⁹ later calculated that a much larger drift (+60° in 100 Myr) was possible for the Earth, assuming a glacial–interglacial variation in oblateness of 1% based on oxygen-isotope data for the Pleistocene glaciations. This estimate was later revised downward (ref. 20) after it was realized that the original calculation had neglected compression of the solid Earth by ice sheets. Subsequent calculations²¹ that included this effect lowered the estimate of the net change in oblateness to ~0.4%. Constrained by this result, subsequent authors^{20,22} have concluded that the change to the Earth's obliquity cannot have been more than 10–20° over the Earth's entire (~450 Myr) glacial history.

These calculations, however, assume that all ice ages were of comparable severity to those in the Pleistocene epoch, which is not necessarily true. Different continental configurations, and possibly lower global-average temperatures, could have resulted in larger $\Delta J_2/J_2$ values for earlier glaciations (see Fig. 1 legend for details of these parameters). A straightforward calculation²³ reveals that if the continents were at one time clustered around one of the poles, as may have occurred during the late Proterozoic²⁴, the change in oblateness from ice loading could have been as high as ~2.6%, provided that the continents were completely covered by ice with a thickness of several kilometres. Under these assumptions, we calculate that the net

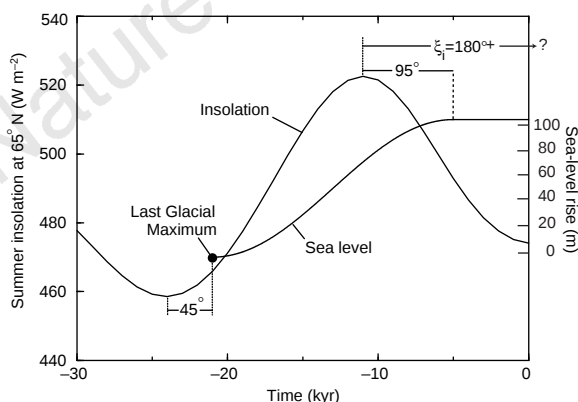


Figure 2 Insolation and relative sea level for the period defining the last deglaciation. The curve showing summer insolation at 65° N was obtained using the orbital solution La90 provided by Laskar *et al.*²⁷ with code to solve the equations of precession and insolation. Last Glacial Maximum took place ~21 kyr ago. The sea-level variation between 21 and 5 kyr is a sinusoidal fit to model results of Peltier³⁴. The slope in the sea-level variation indicates that the initial meltdown was very rapid, and that subsequent changes to the masses of the remaining (Antarctic and Greenland) ice sheets have, since ~5 kyr ago, been negligible. The value of ξ_1 , which is estimated using 26 kyr as the approximate period of the insolation cycle, is ambiguous, as it could be defined either as the time between minimum insolation and glacial maximum or that between maximum insolation and glacial minimum (which itself is difficult to identify precisely).

oblateness variation (with solid-Earth response included) is only ~0.66% on average, owing to differences in insolation cycle amplitude and, thus, maximum ice volume over time. This is still more than 1.5 times the maximum variation thought possible for the Pleistocene epoch, which could have enabled a correspondingly high rate of secular obliquity drift for the late Proterozoic.

The sign and magnitude of the obliquity–oblateness feedback depend on the locations and areas of the continents and on the phase lags ξ_1 and ξ_2 (see Fig. 1 legend). High-latitude ice sheets cause J_2 to decrease because they cancel out part of the Earth's equatorial bulge. For $25^\circ < \xi_1 < 206^\circ$, the resulting secular change in obliquity is positive (see Fig. 1). Low-latitude ice sheets, which might form at times of high obliquity, produce an obliquity drift in the same direction because both the effect on J_2 and the phasing with respect to the obliquity cycle are reversed. Previous studies of climate friction^{13,19,20,22} have assumed that the feedback would be in this direction, based on phase lag estimates for the Pleistocene glaciations. Imbrie *et al.*²⁵ favour $\xi_1 \approx 80^\circ$ (or ~9 kyr) for the 41-kyr obliquity cycle based on cross-spectral analysis of Northern Hemisphere, high-latitude, summer (NHHLS) insolation and $\delta^{18}\text{O}$ values in marine carbonates over the past 2 Myr. However, the actual ages of marine sediments are not known with great precision before ~30 kyr ago, so such inferences are not very firm. A phase lag of $<90^\circ$ is also suggested by analogy with the seasonal cycle, in which the coldest winter temperatures and highest snow accumulation at mid-latitudes occur 1–2 months after winter solstice. Ice volume need not respond in the same manner, though, to the much slower and weaker changes in insolation caused by orbital variations. In colder climates, ice sheets might expand until, or after, NHHLS insolation reaches its peak, causing ice volume to be at least ~180° out of phase with the solar forcing.

Accurate ages are available for the last deglaciation, which is thought to have been triggered by a precessionally induced increase in NHHLS insolation. As Fig. 2 illustrates, the value of ξ_1 for this one event could lie anywhere above 45° , depending on how one interprets the shape of the sea-level (ice-volume) curve. The response of ice volume to NHHLS insolation is evidently not sinusoidal, and presumably depends in a complicated manner on both climate and continental positions. With this in mind, it is quite conceivable that climate friction could act to decrease a planet's obliquity, as required to make Williams's hypothesis work.

To determine what changes to the Earth's obliquity might have been possible, we integrated the equations of motion for the planets

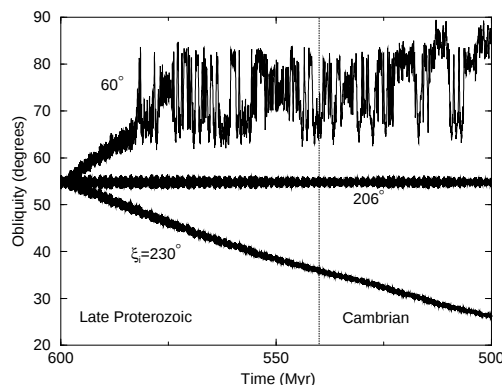


Figure 3 Obliquity 600–500 Myr ago. Shown is the secular drift in obliquity over 100 Myr spanning the late Proterozoic/Cambrian boundary (indicated by a vertical dotted line) for three different values of the ice-sheet-formation phase lag ξ_1 . For the three calculations, $\Delta J_2/J_2 = 0.0262$ and the obliquity is started at 55° , which sets the initial period of the obliquity cycle $P \approx 58.5$ kyr and the phase lag of solid-Earth deformation $\xi_2 \approx 28^\circ$ (ref. 22). The case of positive drift, with $\xi_1 = 60^\circ$, shows the spin axis entering a spin–orbit resonance at an obliquity of ~65° in under 20 Myr; the obliquity then is able to vary chaotically upwards to 90° .

in the Solar System²⁶ and the equations of precession for the Earth²⁷ over 100 Myr using several different values of ξ_i . We assumed, somewhat arbitrarily, that the Earth's obliquity was 55° at 600 Myr ago—the age of the youngest low-latitude glacial deposits^{3–5}. Williams suggested that the obliquity must originally have been >54° because this is the critical latitude above which the poles receive more annually averaged insolation than does the Equator. However, ice sheets are more sensitive to changes in seasonal insolation extremes than to annually averaged insolation, so the Earth's obliquity during Precambrian times need not actually have been this high. Figure 3 shows that under extreme polar ice loading, the Earth's obliquity could have been reduced to within 3° of its present value by 500 Myr ago if $\xi_i = 230^\circ$. (A phase lag near 0° yields a comparable rate of drift in the same direction (Fig. 2)). This result is in good agreement with the geological data referenced earlier¹¹.

Support for the hypothesis that the Earth's obliquity has decreased comes from analysing the orbit of the Moon. The lunar orbit is presently inclined at ~5° to the ecliptic plane. Backwards integrations by Goldreich²⁸ and others²⁹ have predicted that the lunar orbit was inclined by at least 10° to the Earth's equatorial plane when it formed. This result is in apparent conflict with the widely accepted giant-impact theory of lunar origin, in which the Moon accretes from an impact-generated debris disk aligned with the Earth's Equator³⁰. Had the Moon formed in the Earth's equatorial plane, as predicted, its orbit should have aligned itself with the ecliptic plane as the Moon receded from the Earth³¹. The present 5° lunar inclination with respect to the ecliptic plane implies that either the Moon was never in the equatorial plane (that is, that the giant-impact model is wrong) or that it was perturbed away from the equatorial plane early in its history, perhaps as a result of spin-orbit resonances encountered by the Earth and Moon when they were much closer^{31,32}. (The Moon's orbit might also have been inclined by a giant impact, but the extreme size of the necessary impactor²³ (>1,100 km) makes such an event unlikely.)

Alternatively, the lunar orbit may have tilted away from the ecliptic plane more recently in gravitational response to the secular downward drift of the Earth's obliquity around ~600 Myr ago. At present, the Earth's mean obliquity is slowly increasing as a result of tidal interactions with the Moon²⁸. The lunar inclination is decreasing at the same time, so the angular momentum of the Earth–Moon system is approximately conserved. Momentum conservation between the Earth's spin and the Moon's orbit may be approximately written $\sin(\Theta_o/2) = -(L_m/L_\oplus) \sin(\Theta_i/2)$, where $L_\oplus = C\omega$ is the spin angular momentum of the Earth, $L_m = M_m(GM_\oplus)^{1/2} a_m^{1/2}$ is the orbital angular momentum of the Moon, and Θ_o and Θ_i are the angular change to the Earth's obliquity and lunar inclination, respectively, C is one of Earth's principal inertia moments, defined in legend to Fig. 1, and M_m and $M_\oplus$ are the masses of the Moon and the Earth, respectively.

For the late Proterozoic, the distance to the Moon, $a_m \approx 57$ Earth radii, and the rotational velocity of the Earth, $\omega \approx 2\pi/(21 \text{ hours})$ (ref. 33). If we assume that the lunar inclination was originally zero, and that the change that occurred as a result of climate friction was $\Theta_i = +6^\circ$ (allowing for a subsequent 1° reduction, to 5°, as a result of tidal friction), then the required change to the Earth's obliquity is $\Theta_o = -25.4^\circ$. This implies that the Earth's obliquity may have been $23.5^\circ + 25.4^\circ \approx 49^\circ$ for much of the Precambrian period, which is only slightly less than the 54° obliquity suggested by Williams. Given the uncertainty in the actual obliquity that would be required to cause low-latitude glaciation, we believe that the information obtained from the lunar orbit provides additional support for the idea that the low-latitude glaciations in Precambrian times were a consequence of high planetary obliquity. □

Received 12 January; accepted 10 September 1998.

1. Evans, D. A., Beukes, N. J. & Kirschvink, J. L. Low-latitude glaciation in the Palaeoproterozoic era. *Nature* **386**, 262–266 (1997).
2. Williams, G. E., Schmidt, P. W. Paleomagnetism of the Palaeoproterozoic Gowganda and Lorrain formations, Ontario: low paleolatitude for Huronian glaciation. *Earth Planet. Sci. Lett.* **153**, 157–169 (1997).

3. Frakes, L. A. *Climates Throughout Geologic Time* (Elsevier, Amsterdam, 1979).
4. Embleton, B. J. & Williams, G. E. Low paleolatitude of deposition for late Precambrian periglacial varvites in South Australia: implications for palaeoclimatology. *Earth Planet. Sci. Lett.* **79**, 419–430 (1986).
5. Zhang, H. & Zhang, W. Palaeomagnetic data, late Precambrian magnetostratigraphy, and tectonic evolution of eastern China. *Precamb. Res.* **29**, 65–75 (1985).
6. Schmidt, P. W. & Williams, G. E. The Neoproterozoic climatic paradox: Equatorial paleolatitude for Marinoan glaciation near sea level in South Australia. *Earth Planet. Sci. Lett.* **134**, 107–124 (1995).
7. Park, J. K. Paleomagnetic evidence for low-latitude glaciation during deposition of the Neoproterozoic Rapitan Group, Mackenzie Mountains, N.W.T., Canada. *Can. J. Earth Sci.* **34**, 34–49 (1997).
8. Hoffman, P. F., Kaufman, A. J., Halverson, G. P. & Schrag, D. P. A Neoproterozoic snowball Earth. *Science* **281**, 1342–1346 (1998).
9. Meert, J. G. & Van der Voo, R. The Neoproterozoic (1000–540 Ma) glacial intervals: No more snowball earth? *Earth Planet. Sci. Lett.* **123**, 1–13 (1994).
10. Williams, G. E., Schmidt, P. W., Embleton, B. J., Meert, J. G. & Van der Voo, R. The Neoproterozoic (1000–540 Ma) glacial intervals; no more snowball earth?; discussion and reply. *Earth Planet. Sci. Lett.* **131**, 115–125 (1995).
11. Williams, G. E. History of the Earth's obliquity. *Earth Sci. Rev.* **34**, 1–45 (1993).
12. Ward, W. R. Climatic variations on Mars: I. Astronomical theory of insolation. *J. Geophys. Res.* **79**, 3375–3386 (1974).
13. Rubincam, D. P. The obliquity of Mars and “climate friction”. *J. Geophys. Res.* **98**, 10827–10832 (1993).
14. Crowley, T. J. & Baum, S. K. Effect of decreased solar luminosity on Late Precambrian ice extent. *J. Geophys. Res.* **98**, 16723–16732 (1993).
15. Caldeira, K. & Kasting, J. F. Susceptibility of the early Earth to irreversible glaciation caused by carbon dioxide clouds. *Nature* **359**, 226–228 (1992).
16. Vanyo, J. P. & Awramik, S. M. Length of day and obliquity of the ecliptic 850 Ma ago—preliminary results of a stromatolite growth model. *Geophys. Res. Lett.* **9**, 1125–1128 (1982).
17. Awramik, S. M. & Vanyo, J. P. Heliotropism in modern stromatolites. *Science* **231**, 1279–1281 (1986).
18. Rochester, M. G. The secular decrease of obliquity due to dissipative core-mantle coupling. *Geophys. J. R. Astron. Soc.* **46**, 109–126 (1976).
19. Bills, B. G. Obliquity-oblateness feedback: Are climatically sensitive values of obliquity dynamically unstable? *Geophys. Res. Lett.* **21**, 177–180 (1994).
20. Rubincam, D. P. Has climate changed Earth's tilt? *Paleoceanography* **10**, 365–372 (1995).
21. Peltier, W. R. & Jiang, X. Precession constant of the Earth: Variations through the ice-age. *Geophys. Res. Lett.* **21**, 2299–2302 (1994).
22. Ito, T., Masuda, K., Hamano, Y. & Matsui, T. Climate friction: A possible cause for secular drift of Earth's obliquity. *J. Geophys. Res.* **100**, 15147–15161 (1995).
23. Williams, D. M. *The Stability of Habitable Planetary Environments*. Thesis. Pennsylvania State Univ. (1998).
24. Hecht, J. & Scotese, C. R. *Ages of the Earth* (MacMillan, New York, 1997).
25. Imbrie, J. et al. On the structure and origin of major glaciation cycles: 1. linear responses to Milankovitch forcing. *Paleoceanography* **7**, 701–738 (1992).
26. Levison, H. F. & Duncan, M. J. The long-term dynamical behavior of short-period comets. *Icarus* **108**, 18–36 (1994).
27. Laskar, J., Joutel, F. & Robutel, P. Orbital, precessional, and insolation quantities for the Earth from –20 Myr to +10 Myr. *Astron. Astrophys.* **270**, 522–533 (1993).
28. Goldreich, P. History of the lunar orbit. *Rev. Geophys.* **4**, 411–439 (1966).
29. Touma, J. & Wisdom, J. Evolution of the Earth–Moon system. *Astron. J.* **108**, 1943–1961 (1994).
30. Ida, S., Canup, R. M. & Stewart, G. R. Lunar accretion from an impact-generated disk. *Nature* **389**, 353–357 (1997).
31. Rubincam, D. P. Tidal friction and the early history of the Moon's orbit. *J. Geophys. Res.* **80**, 1537–1548 (1975).
32. Touma, J. & Wisdom, J. Resonances in the early evolution of the Earth–Moon system. *Astron. J.* **115**, 1653–1663 (1998).
33. Walker, J. C. G. & Zahnle, K. J. Lunar nodal tide and the distance to the Moon during the Precambrian. *Nature* **320**, 600–602 (1986).
34. Peltier, W. R. Ice age paleotopography. *Science* **265**, 195–201 (1994).

Acknowledgements. We thank H. Levison and M. Duncan for the orbital integration code (SWIFT) and J. Laskar for the code used to integrate the precession equations. We also acknowledge discussions with J. Melosh and R. Canup concerning the lunar inclination. D.M.W. was supported by a NASA graduate student research fellowship awarded in 1995, and J.E.K. was supported by the NASA Exobiology Program.

Correspondence and requests for materials should be addressed to D.M.W. (e-mail: dmw145@psu.edu).

Asymmetric sea-floor spreading caused by ridge–plume interactions

R. Dietmar Müller*, Walter R. Roest† & Jean-Yves Royer‡

* The University of Sydney, School of Geosciences, Building F05, NSW 2006, Australia

† Geological Survey of Canada, 615 Booth Street, Ottawa, ON, K1A 0E9 Canada

‡ Unité Mixte de Recherche 6526, Géosciences Azur, Quai de la Darse - BP 48, 06235 Villefranche sur mer, Cedex, France

Crustal accretion at mid-ocean ridges is generally modelled as a symmetric process. Regional analyses, however, often show either small-scale asymmetries, which vary rapidly between individual spreading corridors, or large-scale asymmetries represented by consistent excess accretion on one of the two separating plates over geological time spans^{1–6}. In neither case is the origin of the

asymmetry well understood. Here we present a comprehensive analysis of the asymmetry of crustal accretion over the past 83 Myr based on a set of self-consistent digital isochrons⁷ and models of absolute plate motion^{8,9}. We find that deficits in crustal accretion occur mainly on ridge flanks overlying one or several hotspots. We therefore propose that asymmetric accretion is caused by ridge propagation towards mantle plumes or minor

ridge jumps sustained by asthenospheric flow^{10,11} between ridges and plumes. Quantifying the asymmetry of crustal accretion provides a complementary approach to that based on geochemical¹² and other geophysical data^{13,14} in helping to unravel how mantle plumes and mid-ocean ridges are linked through mantle convection processes.

To analyse global spreading asymmetries, we constructed a grid of

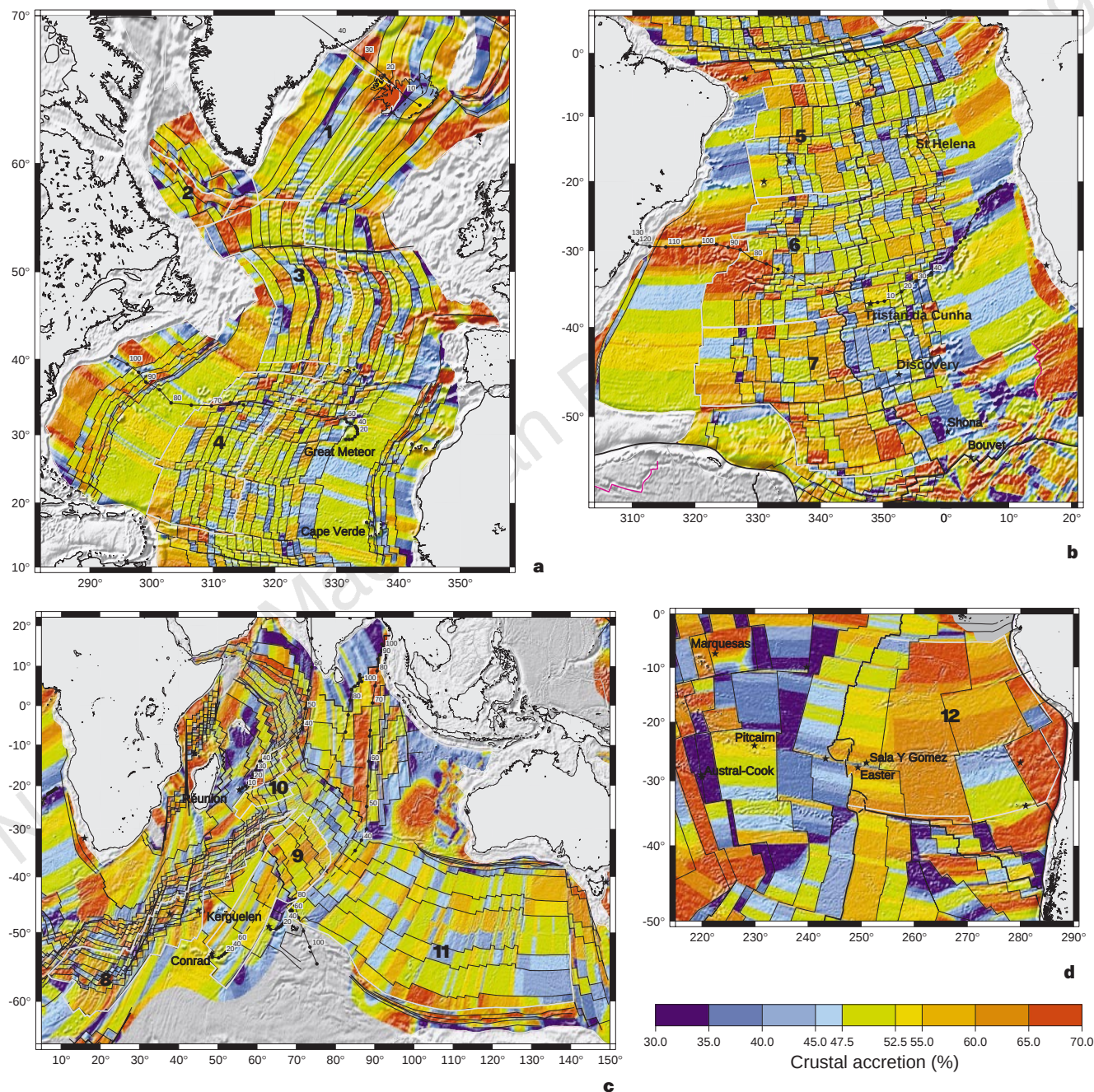


Figure 1 Crustal accretion rates of conjugate plates in four areas, illuminated by marine gravity anomalies²⁷. **a**, The North Atlantic; **b**, the South Atlantic; **c**, the Indian Ocean; and **d**, the South Pacific Ocean. The isochrons shown are 5 (10.9 Myr ago), 6 (20.1 Myr ago), 13 (33.1 Myr), 18 (40.1 Myr), 21 (47.9 Myr), 25 (55.9 Myr), 31 (67.7 Myr), 34 (83.5 Myr), M0, M4, M10, M16, M21 and M25 (ref. 7). Isochrons in the southwest Pacific are based on Munschy *et al.*²⁸. Areas for which we analysed the asymmetry of crustal accretion through time are outlined by grey polygons and numbered from 1 to 12. Also shown are synthetic tracks of hotspots crossing the ridge (from ref. 8); labels are in million of years (see ref. 29 for a discussion on model accuracy). The gaps in hotspot tracks created by ridge

crossings are dashed. Hotspots that have been close to a ridge for long geological time periods may simultaneously produce hotspot tracks on two plates (for example, Kerguelen, Tristan). Present-day locations of hotspots are shown as stars, and the main hotspots are labelled. Active plate boundaries are shown as bold black lines, whereas extinct ridges are coloured magenta. Continental margins are light grey, and areas of insufficient data coverage are dark grey; AAD indicates Australia-Antarctic discordance. Asymmetries off western Australia represent deviations from spreading rates based on best-fit half-stage poles, as conjugate Indian Ocean floor is subducted.

crustal accretion rates using the global age data set of Müller *et al.*⁷ (Fig. 1). Resolution is limited by the isochron spacing and the grid node interval of 0.1° . We assessed the internal consistency by extracting the spreading asymmetry for conjugate plate pairs from two selected areas of equal time range for Africa–South America (area 5, Fig. 1b) and Australia–Antarctica (area 11, Fig. 1c), and tested whether the percentages of crustal accretion add up to 100%. The mean error for these two plate pairs is $0.23 \pm 1.8\%$. However, errors in cumulative plate accretion for the same plate pairs are much smaller, that is, 0.075% for area 5 and 0.005% for area 11.

At relatively small scales, asymmetries in crustal accretion vary between individual spreading corridors and through time without clear trends, and independently of the presence of ridge propagators (Fig. 1). This is in accordance with patterns of spreading asymmetry found on a local scale on the Mid-Atlantic Ridge^{3,5,6}. Such short-term asymmetries in crustal accretion may be caused by long-lived asymmetries in small-scale convection under the ridge¹⁵, by intra-segment ridge propagation³ or variations in melt supply⁵. On the timescales that we can resolve (stage durations of ~ 6 – 14 Myr) there is no correlation between the short-term asymmetry of spreading and half-spreading rates, or the migration of the ridge relative to the mantle.

In order to analyse short- versus long-term spreading asymmetries (Fig. 2), we calculated the rates of mid-ocean-ridge migration

relative to the mantle, parallel to the spreading direction, using absolute plate motion models^{8,9} (Fig. 3). Small-scale, short-term asymmetries do not correlate with spreading rates or absolute ridge migration rates within the resolution of our data. However, Figs 2 and 4 clearly show cumulative, long-term asymmetries. We first evaluate their correlation with ridge migration rates. The South Atlantic ridge (areas 5, 6, 7) exhibits the slowest rates of ridge migration of the ridges considered here (2.4 – 3.9 mm yr⁻¹). Rates of excess accretion have been consistently high on South America (2.4 – 4.1%). The ridges in the North Atlantic (areas 1, 2, 3, 4) all migrated with average rates of 16 – 19 mm yr⁻¹, and show small totals of excess accretion between 0.6% and 0.7% on the leading plate. In contrast, the southwest (area 8) and western southeast Indian (area 9) ridges migrated northwards at average rates of ~ 16 and 49 mm yr⁻¹, but show excess accretion on the trailing plate (that is, Antarctica) of 3.1% and 1.7% , respectively. The central Indian ridge (area 10) migrated northeastwards about as fast as the western southeast Indian ridge, but shows a 'flip' in excess accretion from the leading plate to the trailing plate at chron 13 (33.1 Myr ago), resulting in negligible cumulative asymmetry (0.5%). On the eastern southeast Indian ridge (area 11) the leading Australian flank shows excess accretion on average for the past 35 Myr, while the ridge has been moving north at rapid rates averaging ~ 40 mm yr⁻¹

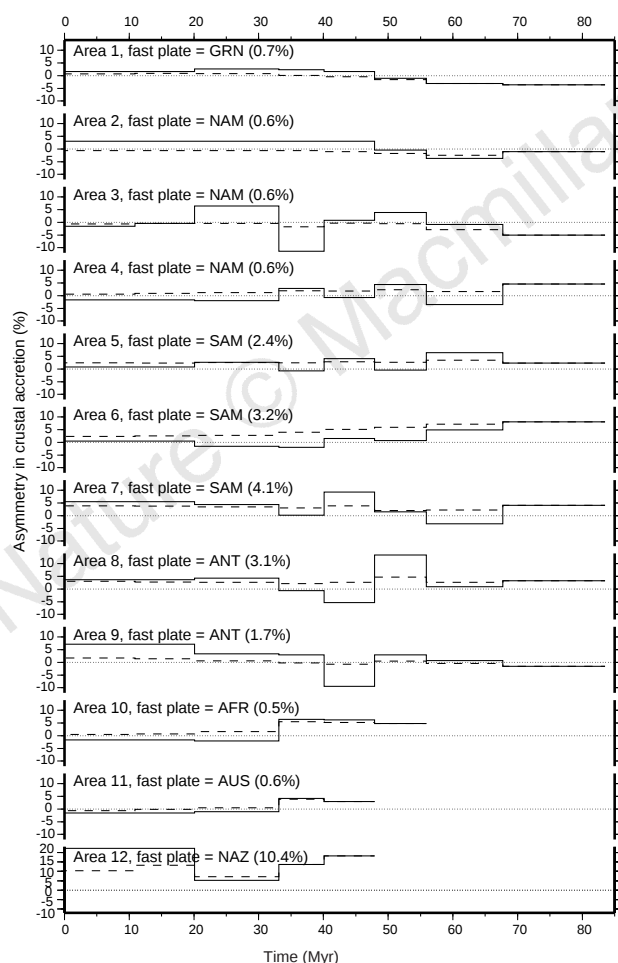


Figure 2 Spreading asymmetry by stage (continuous lines) versus time for areas marked in Fig. 1. The total, cumulative difference in accreted area for each plate pair is given in per cent (dashed lines). For each plate pair, the two graphs are identical for the first stage (83.5–67.7 Myr), and then diverge to express fluctuations in accretion asymmetry through time versus cumulative trends. The zero-age value of the cumulative curve distinguishes 'fast plates' (cumulative excess accretion) from 'slow plates'. GRN, Greenland; NAM, North America; SAM, South America; ANT, Antarctica; AFR, Africa; AUS, Australia; NAZ, Nazca.

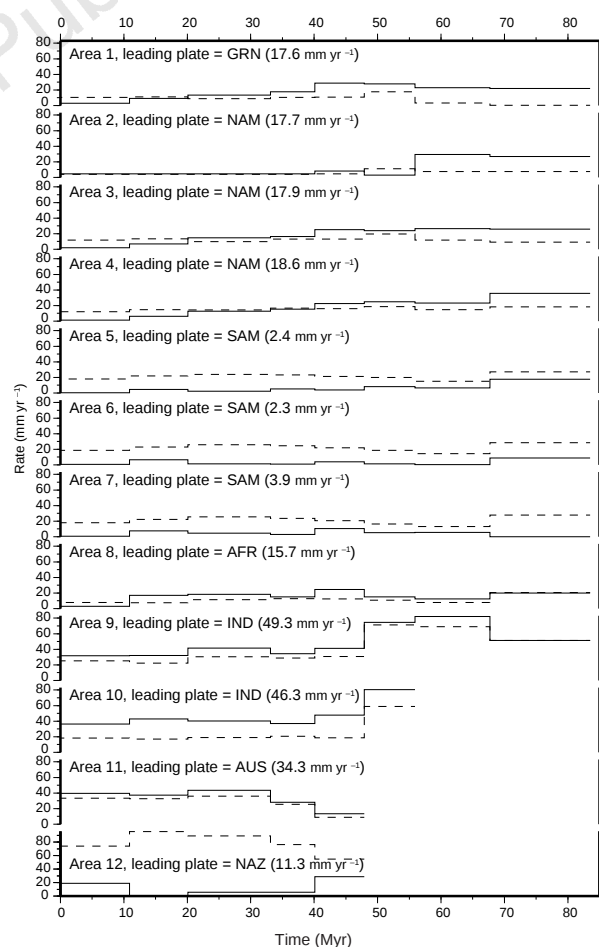


Figure 3 Rate of the ridge-normal component of the migration of mid-ocean ridges relative to the mantle by stage versus time (solid lines). Rates are computed for points roughly in the middle of the mid-ocean-ridge segments analysed (Fig. 1). Half-spreading rates are shown as dashed lines. The average, cumulative ridge migration rate is given for the leading plates, that is, the plates which move faster relative to the mantle than their conjugates. IND, India; other abbreviations as in Fig. 2.

since 33 Myr ago. On the East Pacific Rise (area 12, Fig. 1d) ridge migration is quite slow (11.3 mm yr^{-1}) despite fast spreading rates (Fig. 3). Here we find excess accretion on the leading Nazca plate (10.4% since 40 Myr ago).

Thus, contrary to the model suggested by Stein *et al.*¹⁶, there appears to be no correlation between long-term spreading asymmetry and migration of the spreading ridge over a fixed or slowly moving mantle. Based on a fluid mechanical flow model, Stein *et al.*¹⁶ suggested that trailing plates should accrete faster. However, Sleep¹⁷ and Fujita and Sleep¹⁸ showed that the fast side of a spreading ridge is hotter and weaker, and therefore more susceptible to dyke intrusion. This acts as a stabilizing mechanism, and may help explain why ridge migration does not necessarily result in long-term asymmetries. Only the spreading ridge between India and the trailing, Antarctic, ridge flank shows consistent excess accretion on the trailing plate as implied by Stein *et al.*¹⁶. This ridge also moved at relatively high rates of $30\text{--}80 \text{ mm yr}^{-1}$ relative to the mantle since chron 34 (83 Myr ago) (Fig. 3), suggesting a minimum 'absolute' ridge migration speed for Stein *et al.*'s¹⁶ mechanism to have an effect on the asymmetry of spreading.

A connection between asymmetries in crustal accretion and in asymmetric depth-age behaviour as suggested by Hayes⁴ is not evident. For instance, in the southern South Atlantic (area 7) there is excess accretion on the anomalously deep¹⁹ South American ridge flank, whereas in the western southeast Indian Ocean (area 9) pronounced excess accretion occurs on the anomalously shallow²⁰ Antarctic ridge flank. Where plate-wide asymmetries are observed, they do not occur in all spreading corridors (Fig. 1). Stein *et al.*¹⁶ noted that if hotspot-ridge asthenospheric flow is relatively fast compared to ridge migration speed, then the latter alone will not predict spreading asymmetry.

The generation of long-term asymmetries requires dyke intrusion to consistently move towards the slower-spreading flank, which is cooler in the absence of hotspots. To investigate the effect of hotspots on spreading asymmetries, we show synthetic tracks of

main hotspots⁸ (Fig. 1). Hotspots may cause ridge jumps by pressure-release melting of material flowing towards the ridge, causing excess dyke intrusion on the ridge flank proximal to the hotspot. The buoyancy of the plume material increases the spreading force, which depends on the buoyancy times its depth^{13,18}. Deep plume buoyancy of sufficient magnitude can cause the point of maximum tension to be located off-axis, close to the plume.

A clear example of such 'trapping' of a ridge by a hotspot is seen in the South Atlantic (area 6). Here successive ridge jumps towards the Tristan da Cunha hotspot just east of the ridge caused consistent excess accretion on the South American plate (Fig. 1b). The absence of hotspots close to the mid-ocean ridge since initiation of spreading explains why spreading asymmetry is negligible in the Labrador Sea, in most of the North Atlantic and in the southeast Indian Ocean (area 11). In contrast, deficits in crustal accretion are found on those ridge flanks that are underlain by hotspots in the Norwegian-Greenland Sea, the South Atlantic and in the southwest and central Indian oceans. The current locus of active volcanism on Iceland is located on eastern Iceland²¹, indicating recent ridge jumps to the east, whereas earlier jumps were dominantly to the west (Fig. 1a). This supports models of a stationary Iceland hotspot²² with a vertical cylindrical structure²³ rather than a laterally moving plume over which the ridge has migrated from west to east. In the South Atlantic, the St Helena, Tristan da Cunha, Discovery, Shona and Bouvet hotspots are all located on the African plate, and have caused both large-scale ridge propagation events as well as successive jumps of the ridge axis towards Africa, leaving excess crust on South America. The same relationship of net asymmetries with hotspots on the slowly spreading ridge flank are found for spreading between Africa and India (area 10), showing a deficit in accretion on the African plate during the past 33 Myr, when Réunion was located underneath the African plate, and a deficit on the Indian plate before 33 Myr ago when the hotspot was situated east of the ridge (see also ref. 12). Kerguelen appears to have channelled asthenospheric material towards the southeast Indian ridge^{19,20}, resulting in crustal deficits in the corridors proximal to the hotspot on Antarctica (Fig. 1c), even though the plume is located $\sim 1,000 \text{ km}$ away from the ridge (see also ref. 12). The most extreme asymmetries observed, on the East Pacific Rise, are expressed as deficits in crustal accretion on the French Polynesian part of the Pacific plate, which is underlain by a thermal²⁴ or dynamically supported²⁵ 'superswell' and numerous hotspots (Fig. 1d), indicating that successive ridge propagators/jumps²⁶ may have been triggered by plume-ridge interaction.

Fig. 1a-d illustrates that spreading asymmetry is ubiquitous in all ocean basins, but its occurrence is not dependent on the presence of ridge propagators¹⁴ (Fig. 1). This indicates that discrete ridge jumps, such as those described by Weiland *et al.*³, play an important role in causing cumulative asymmetries in accretion. In the absence of medium-to-large hotspots which channel material towards the ridge, small-scale ridge jumps occur randomly, with little net accretion asymmetry through time. Ridge-hotspot interaction biases this process by causing successive ridge jumps towards the plume. We conclude that ridge-plume interaction is the main cause of spreading asymmetries, even for large ridge-plume distances. □

Received 8 May; accepted 4 August 1998.

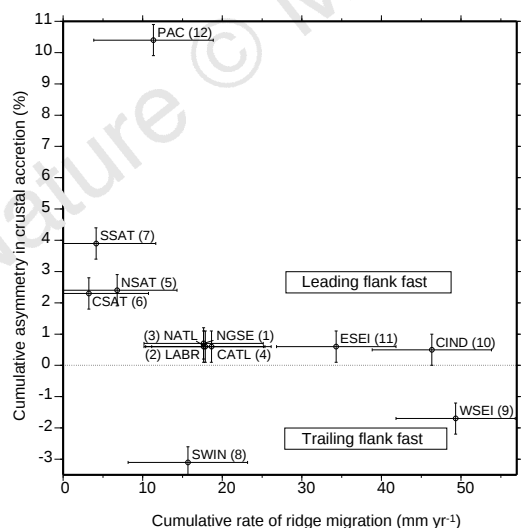


Figure 4 Cumulative rate of ridge migration versus spreading asymmetry and estimated errors. Data are given for areas marked in Fig. 1 averaged for time periods shown in Figs 2 and 3. Most leading ridge flanks are fast, because the conjugate trailing flanks are underlain by one or several mantle plumes. Exceptions are the western southeast Indian ridge (area 9), where the trailing flank is fast, in accordance with Stein's *et al.*'s¹⁶ model, and the southwest Indian ridge (area 8), which is very complex, and has experienced severe changes in spreading direction. PAC, South Pacific; SSAT, southern South Atlantic; CSAT, central South Atlantic; NSAT, northern South Atlantic; NATL, North Atlantic; CATL, central North Atlantic; LABR, Labrador Sea; NGSE, Norwegian Greenland Sea; ESEI, eastern southeast Indian ridge; WSEI, western southeast Indian ridge; SWIN, southwest Indian ridge; CIND, central Indian ridge.

- Stoddard, P. R. & Stein, S. A kinematic model of ridge-transform geometry evolution. *Mar. Geophys. Res.* **10**, 181–190 (1988).
- Cande, S. C., LaBrecque, J. L. & Haxby, W. F. A high resolution seafloor spreading history of the South Atlantic. *J. Geophys. Res.* **93**, 13479–13492 (1988).
- Weiland, C., Wilson, D. S. & MacDonald, K. High-resolution plate reconstruction of the southern Mid-Atlantic Ridge. *Mar. Geophys. Res.* **17**, 143–166 (1995).
- Hayes, D. E. Nature and implications of asymmetric seafloor spreading. *Geol. Soc. Am. Bull.* **87**, 994–1002 (1976).
- Sempéré, J.-C. *et al.* The Mid-Atlantic Ridge between 29°N and 31°30'N in the last 10 Ma. *Earth Planet. Sci. Lett.* **130**, 45–55 (1995).
- Kleinrock, M. C. & Humphreys, S. E. Structural asymmetry of the TAG rift valley: Evidence from a near-bottom survey for episodic spreading. *Geophys. Res. Lett.* **23**, 3439–3442 (1996).
- Müller, R. D., Roest, W. R., Royer, J.-Y., Gahagan, L. M. & Sclater, J. G. Digital isochrons of the world's ocean floor. *J. Geophys. Res.* **102**, 3211–3214 (1997).
- Müller, R. D., Royer, J.-Y. & Lawver, L. A. Revised plate motions relative to the hotspots from combined Atlantic and Indian Ocean hotspot tracks. *Geology* **21**, 275–278 (1993).

9. Müller, R. D., Royer, J.-Y. & Lawver, L. A. *A Review of Absolute Plate Motion Models from Jurassic to Present Day* (Tech. Rep. No. 112) (Institute for Geophysics, Univ. Texas, Austin, 1991).
10. Morgan, W. J. in *The Oceanic Lithosphere* (ed. Emiliani, C.) 443–487 (Wiley, New York, 1981).
11. Phipps-Morgan, J., Morgan, W. J., Zang, Y.-S. & Smith, W. H. F. Observational hints for a plume-fed, suboceanic asthenosphere and its role in mantle convection. *J. Geophys. Res.* **100**, 12753–12767 (1995).
12. Schilling, J. G. Fluxes and excess temperatures of mantle plumes inferred from their interaction with migrating mid-ocean ridges. *Nature* **352**, 397–403 (1991).
13. Sleep, N. H. Hotspots and mantle plumes: Some phenomenology. *J. Geophys. Res.* **95**, 6715–6736 (1990).
14. Small, C. Observations of ridge-hotspot interactions in the Southern Ocean. *J. Geophys. Res.* **100**, 17931–17946 (1995).
15. Rouzo, S., Rabinowicz, M. & Briais, A. Segmentation of mid-ocean ridges with an axial valley induced by small-scale mantle convection. *Nature* **74**, 795–798 (1995).
16. Stein, S., Melosh, H. J. & Minster, J. B. Ridge migration and asymmetric sea-floor spreading. *Earth Planet. Sci. Lett.* **36**, 51–62 (1977).
17. Sleep, N. H. Formation of oceanic crust: some thermal constraints. *J. Geophys. Res.* **80**, 4032–4042 (1975).
18. Fujita, K. & Sleep, N. H. Membrane stresses near mid-ocean ridge-transform intersections. *Tectonophysics* **50**, 207–221 (1978).
19. Phipps-Morgan, J. & Smith, W. H. F. Flattening of the sea-floor depth-age curve as a response to asthenospheric flow. *Nature* **359**, 524–527 (1992).
20. Grevemeyer, I. Hotspot-ridge interaction in the Indian Ocean: constraints from Geosat/ERM altimetry. *Geophys. J. Int.* **126**, 796–804 (1996).
21. Oskarsson, N., Steinthorsson, S. & Sigvaldason, G. E. Iceland geochemical anomaly: origin, volcanotectonics, chemical fractionation and isotope evolution of the crust. *J. Geophys. Res.* **90**, 10011–10025 (1985).
22. Lawver, L. A. & Müller, R. D. Iceland hotspot track. *Geology* **22**, 311–314 (1994).
23. Wolfe, C. J., Bjarnason, I. T., VanDecar, J. C. & Solomon, S. Seismic structure of the Iceland mantle plume. *Nature* **385**, 245–247 (1997).
24. McNutt, M. K. & Judge, A. V. The superswell and mantle dynamics beneath the South Pacific. *Science* **248**, 933–1048 (1990).
25. Stein, C. A. & Stein, S. in *The Mesozoic Pacific: Geology, Tectonics, and Volcanism* (eds Pringle, M. S., Sager, W. W., Slater, W. V. & Stein, S.) 53–76 (Am. Geophys. Union, Washington DC, 1993).
26. Goff, J. A. & Cochran, J. R. The Bauer scarp ridge jump: A complex tectonic sequence revealed in satellite altimetry. *Earth Planet. Sci. Lett.* **141**, 21–33 (1996).
27. Sandwell, D. T. & Smith, W. H. F. Marine gravity anomaly from ERS-1, Geosat and satellite altimetry. *J. Geophys. Res.* **102**, 10039–10045 (1997).
28. Munsch, M., Antoine, C. & Gachon, A. Evolution tectonique de la région Tuamotu, océan Pacifique central. *C.R. Acad. Sci. Paris Ser. IIa* **323**, 941–948 (1996).
29. Müller, R. D., Royer, J.-Y. & Lawver, L. A. Revised plate motions relative to the hotspots from combined Atlantic and Indian hotspot tracks: reply. *Geology* **22**, 277–278 (1994).

Acknowledgements. We thank N. Sleep and S. Stein for comments on the manuscript, and R. Buick and A. Dutkiewicz for discussions.

Correspondence and requests for materials should be addressed to R.D.M. (e-mail: dietmar@es.su.oz.au).

Implications of *Deltatheridium* specimens for early marsupial history

Guillermo W. Rougier^{*,†}, John R. Wible^{‡,*} & Michael J. Novacek[†]

^{*} Department of Anatomical Sciences and Neurobiology, School of Medicine, University of Louisville, Louisville, Kentucky 40292, USA

[†] Department of Vertebrate Paleontology, American Museum of Natural History, New York, New York 10024, USA

[‡] Section of Mammals, Carnegie Museum of Natural History, Pittsburgh, Pennsylvania 15206, USA

We describe here two new specimens of the mammal *Deltatheridium pretrituberculare* from the Late Cretaceous period of Mongolia. These specimens provide information on tooth replacement in basal therian mammals and on lower jaw and basicranial morphology. Deltatheroidans, known previously from isolated teeth, partial rostra and jaws from the late Cretaceous of Asia^{1–4} and possibly North America^{5,6}, have been identified variously as eutherians^{1,7,8}, as basal metatherians (the stem-based clade formed by marsupials and their extinct relatives)^{3,9–11}, or as an outgroup to both eutherians and metatherians^{2,12–15}. Resolution of these conflicting hypotheses and understanding of the early evolution of the therian lineage have been hampered by a sparse fossil record for basal therians. The new evidence supports metatherian affinities for deltatheroidans and allows a comprehensive phylogenetic analysis of basal metatherians and marsupials. The presence of specialized marsupial patterns of tooth replacement and cranial

vascularization in *Deltatheridium* and the basal phylogenetic position of this taxon indicate that these features are characteristic of Metatheria as a whole. Other morphological transformations recognized here secure the previously elusive diagnosis of Metatheria^{3,14,15}. The new specimens of *Deltatheridium* illustrate the effectiveness of fairly complete fossil specimens in determining the nature of early evolutionary events.

Two specimens of *Deltatheridium pretrituberculare* were recovered from Ukhaa Tolgod, the Mongolian Late Cretaceous locality notable for the abundance and exquisite preservation of dinosaurs, lizards, birds and mammals¹⁶. The first specimen, PSS-MAE 133, is an adult represented by a partial skull, complete lower jaws and two ulnae. The second, PSS-MAE 132, is a juvenile, consisting of nearly complete jaws, disarticulated skull bones and several postcranial elements. Comparison of all known specimens of *Deltatheridium*^{1,2} indicate that there is a single species, *D. pretrituberculare*, with no subspecies (contra ref. 2).

The new specimens of *Deltatheridium* provide the following, revised dental formula: 14/3 C1/1 P3/3 M4/4. In contrast to previous observations^{1,2,8,13}, three lower incisors and four upper molars are present. I₂, the largest incisor, is 'staggered' as in several groups of metatherians¹⁷. I₂ and I₃ are spatulated, but the crown of I₁ is missing in all known specimens. Previous reports of either one or two incisors in *Deltatheridium* (the main feature distinguishing two reported subspecies²) are artefactual. A tiny fourth upper molar is present bilaterally in PSS-MAE 133 (Fig. 1); this tooth and its alveolus were missing or unrecognized in earlier specimens, which led to the incorrect count of three upper molars. The presence of only three molars is unusual among metatherians and was used as evidence against the possible marsupial affinities of *Deltatheridium*^{2,12,13}. M⁴ is positioned lingually to the metacone of M³, so it continues the lingual curvature and reduction of the posterior molar dentition present in deltatheroidans, *Holoclemensia* and *Potamotelses*¹⁸.

The skull of *Deltatheridium* also has derived features indicating marsupial affinity, including a premaxilla with a posteriorly directed process that reaches the alveolus of the canine (Fig. 1), a feature widely present in metatherians but absent in basal eutherians and Mesozoic outgroups. The lower jaw has a distinctive shelf-like, medially inflected angle (Fig. 2), a metatherian feature that is absent in Mesozoic outgroups, monotremes and eutherians¹⁹.

Study of the second specimen, PSS-MAE 132, indicates that the braincase morphology of *Deltatheridium* was similar to that of therian mammals in genera, with an extensive squama of the squamosal and an anterior lamina absent from the braincase wall, and of marsupials in particular, with a deep zygomatic process of the squamosal, also present in the Early Cretaceous prototribosphenidan *Vincelestes*. The petrosals (Fig. 3) are similar to those attributed to metatherians from the Late Cretaceous of North America²⁰, showing two major metatherian synapomorphies: first, an absence of vascular sulci on the promontorium, indicating a marked reduction or complete absence of the stapediaal arterial system; and second, a small, horizontally orientated prootic canal connected to the postglenoid venous system.

Perhaps the most distinctive similarity between *Deltatheridium* and living marsupials is the tooth-replacement pattern. Unlike any other group of mammals, marsupials replace only one tooth postnatally, the last premolar^{21–23}. The jaws of PSS-MAE 132 show several teeth in the process of eruption (Fig. 4). Although the first two premolars show substantial wear and are fully erupted, the canine and P₃ are erupting; the former is about halfway out and the posterior accessory cusp of the latter is at the level of the alveolar margin. M₃ has a fully functional trigonid, but the talonid is still at the level of the alveolar margin, whereas the smaller M₄ is lodged deep in a crypt and rotated about 50° from the horizontal axis and 35° from the mandibular axis. Regular X-rays, computerized-tomography scanning and dissection of the right jaw did not

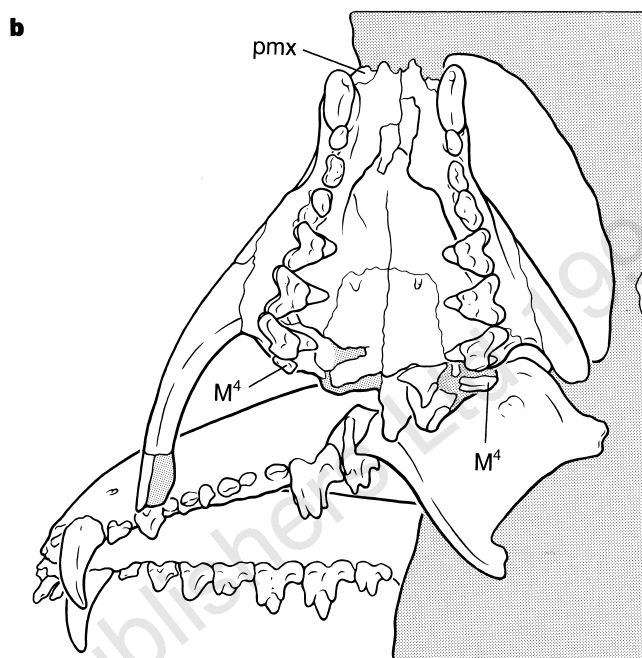


Figure 1 *Deltatheridium pretrituberculare* specimen PSS-MAE 133, with palate in occlusal view and mandible in oblique dorsal view. **a**, Photograph. **b**, Diagram. Four upper molars are preserved in each side of the maxilla. The right M^4 is in place but the left has fallen out of the alveolus and is floating near M^3 . M^4 is a very small, biradicated tooth; the metastylar portion of the crown is markedly reduced and only the greatly reduced anterior half of the stylar shelf remains. No parastyle (cusp A) is present. A small acuspate stylar crest borders labially the anterior portion of the stylar shelf. The paracone is the only functional cusp,

centrally placed in the minute crown. A small metacone lies on the distal slope of the paracone. The protocone is lingually reduced to a small, unbasined shelf. The M^4 lingual root supports the protoconal area and the paracone/metacone, whereas the labial root supports the stylar shelf remnant. The palate lacks palatal vacuities, and there is a large dorsal process of the palatine in the orbit. This process extends anterodorsally towards the infraorbital foramen and encloses the sphenopalatine foramen. Pmx, premaxilla.

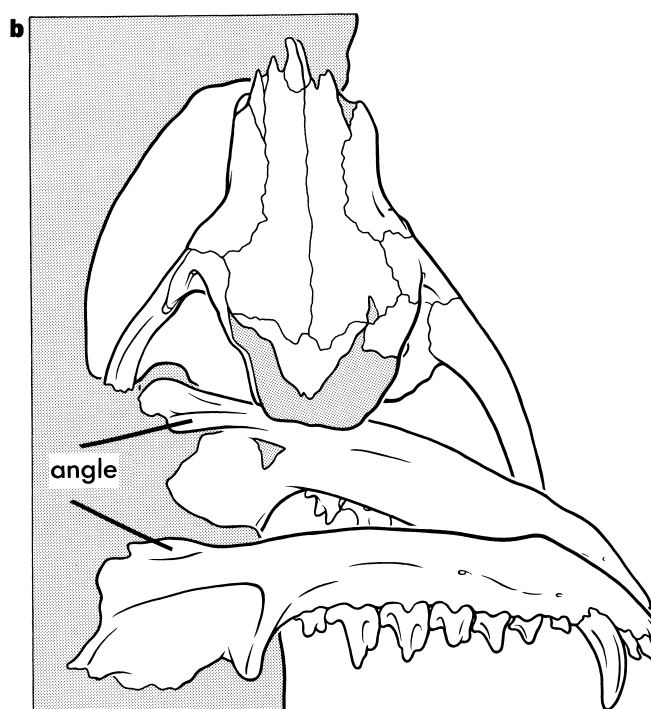


Figure 2 *Deltatheridium pretrituberculare* PSS-MAE 133, with rostrum in dorsal view and mandible in oblique ventral view. **a**, Photograph. **b**, Diagram. The shelf-

like, medially inflected angle of the mandible resembles that in marsupials¹⁹.

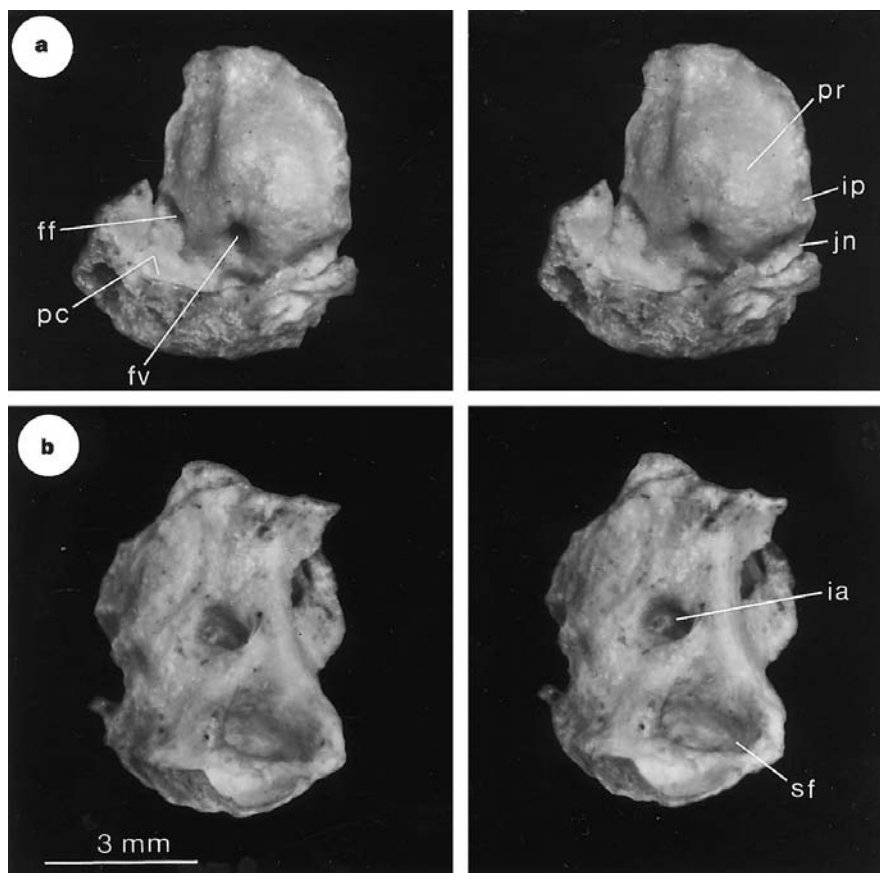


Figure 3 Stereophotograph of the petrosals of *Deltatheridium pretrituberculare* specimen PSS-MAE 132. **a**, Right petrosal in ventral view. **b**, Inverted left petrosal in endocranial view. These petrosals are from a juvenile specimen represented by a partial skeleton including, in addition to petrosals, squamosal, parieto-supraoccipital, lower jaws and several isolated postcranial elements. **a**, Ventrally, the petrosal has a deep post-promontorial recess and a tall, laminar caudal tympanic process that extends towards the jugular notch, as in marsupials. The

small, horizontally orientated prootic canal faces anteriorly and is hidden from view (location indicated by angled line). **b**, Endocranially, a proportionally large floor for the trigeminal ganglion, a primitive feature for Theria, is seen. ff, facial foramen; fv, fenestra vestibuli; ia, internal acoustic meatus; ip, inferior petrosal sinus notch; jn, jugular notch; pc, postotic canal; pr, promontorium; sf, subarcuate fossa.

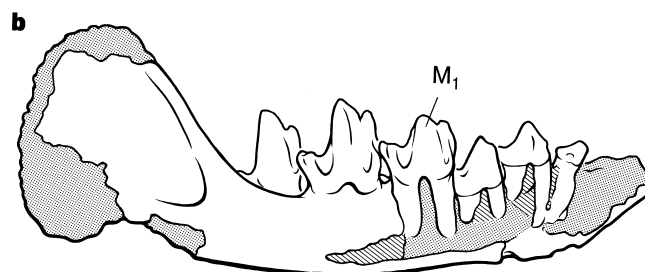


Figure 4 The right lower jaw of *Deltatheridium pretrituberculare* PSS-MAE 132. **a**, Photograph. **b**, Diagram. M_1 , first lower molar. M_3 is placed lingually to the rest of the tooth row, and the talonid of M_2 is still not locked with the anterior accessory cusps of M_3 as it is in adults. The molars rotate labially and slightly anteriorly

during eruption, as among living didelphids. The erupting canine is preserved only in the left jaw. The last deciduous premolar in therians, and (where known) in Mesozoic mammals, is molariform^{21,25}, as is probably the case for the deciduous last premolar in deltatheroidans.

reveal any evidence of replacement teeth. This indicates that the distinctive marsupial tooth-replacement pattern, in which only the deciduous P^3/P_3 are replaced and the deciduous P^1P^2/P_1P_2 are retained throughout life, was also present in *Deltatheridium*. A similar pattern has been described in the fossil metatherian *Alphadon*²².

Although the nature of the deltatheroidan replacement pattern may resemble that in marsupials and *Alphadon*, the timing of replacement is different. In marsupials, represented by fossil and living didelphids, some dasyurids and probably *Alphadon* (based

on a single late juvenile jaw), P_3 erupts after or at about the same time as M_4 , at which point the canine is also fully erupted^{21–23}. The onset of the replacement in *Deltatheridium* occurs relatively earlier, before M_3 , M_4 and canines are completely erupted. Eutherians²⁴, deltatheroidans and other Mesozoic groups²⁵ exhibit what is thought to be the primitive condition, in which the premolars are replaced before all molars are in place. The suppression of replacement of most of the antemolar dentition has been proposed as an adaptation related to the extended period of 'nipple fixation' present in marsupials²¹, during which time the neonate attaches itself to the mother for a period longer than gestation to complete development²⁶. If the metatherian patterns of skull development²⁷, tooth replacement and reproduction are correlated²¹, deltatheroidans may have already possessed the basic marsupial reproductive pattern.

This new anatomical information about *Deltatheridium* allows the first comprehensive study of the early evolution and relationships of metatherians and basal tribosphenic mammals. This analysis provides (Fig. 5) phylogenetic resolution at the base of Tribosphenida and establishes deltatheroidans as basal metatherians, and possibly as the most basal group, depending on the position of *Holoclemensia*. The inclusion of deltatheroidans in Metatheria is well supported by 18 craniodental synapomorphies, including the suppression of the replacement of the anterior premolars (Fig. 5). Information about a fairly complete basal metatherian, such as *Deltatheridium*, can be integrated with the growing morphological and molecular database for marsupials^{10,28,29}. This integration is essential for understanding the early radiation of not only marsupials but also eutherians. Although not strongly supported, a major dichotomy within post-deltatheroidan metatherians is shown in the cladogram (Fig. 5) indicating the presence of a North American–Asiatic clade and a South American–Australian one, corresponding roughly to Boreometatheria and Notometatheria²⁸. These results are in agreement with Cretaceous/Early Cenozoic biogeography, indicating a possible South American origin for the common ancestor of all living marsupials and the subsequent dispersal of this ancestor to Australia^{11,29}.

The Late Cretaceous deltatheroidans are basal metatherians, but they are probably removed from the common ancestor of all metatherians, as indicated by the many synapomorphies shared with later metatherians and marsupials. Despite extrapolations, based on calculations of gene-mutation rates, that indicate a Mid-Cretaceous origin for many modern lineages of marsupials and placentals³⁰, our phylogenetic analysis shows that no members of the crown group Marsupalia have been found yet in the Mesozoic. □

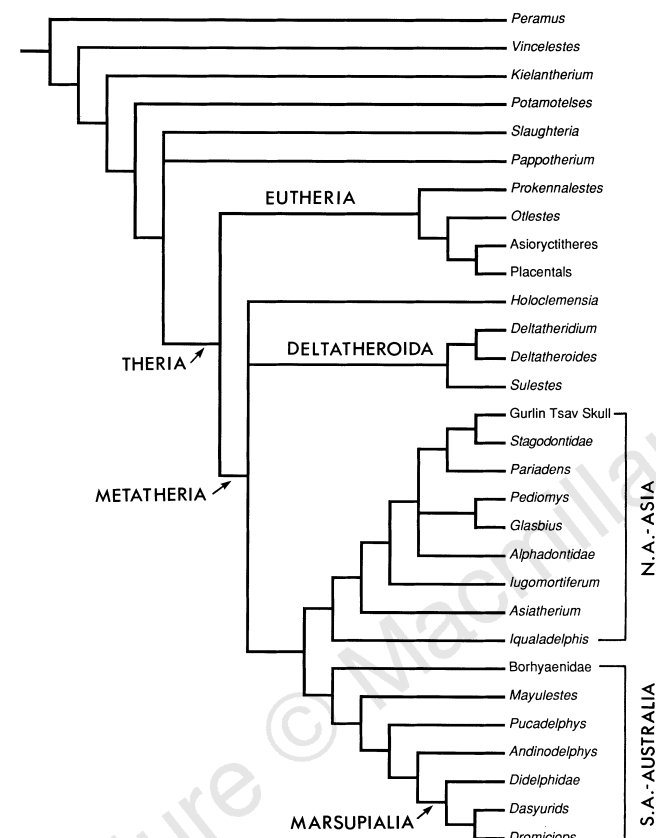


Figure 5 Simplified cladogram based on the consensus tree of 144 equally most parsimonious trees. Trees were obtained by 150 replications of heuristic searches using PAUP with a database of 156 craniodental characters, representing 365 morphological transformations across 48 taxa. Six of these taxa are represented by only one tooth and were subsequently deleted from the study because of their incompleteness. Tree length of the individual trees is 570; consistency index = 0.444; retention index = 0.663. Within Metatheria, the South American–Australian clade and the North American–Asiatic clade were not very stable, and portions of them collapsed in slightly less parsimonious trees. In many of those slightly longer trees, the 'carnivorous' metatherian lineages, such as borhyaenids, the Gurlin Tsav specimen and stagonodontids, formed a paraphyletic series at the base of the post-deltatheroidan metatherians. Diagnostic features of Metatheria are: three premolars; seven postcanine tooth families; single rooted upper canine; procumbent first upper premolar separated by diastema from P^2 ; deep ectoflexus on the penultimate and preceding molars; metacone slightly smaller than paracone; last lower molar rotated during eruption; deciduous canine absent; replacement of deciduous P^1/P_1 and deciduous P^2/P_2 absent; labial mandibular foramen absent; 'Meckelian' groove absent; 'coronoid' facet absent; palatal process of premaxilla approaches or reaches canine alveolus; sulcus for anterior distributary of transverse sinus posterolateral to subarcuate fossa; foramen for ramus superior of the stapedial artery absent; sulcus for stapedial artery absent; jugular foramen separated from inferior petrosal sinus opening; ascending canal absent. See Supplementary information for the full tree, diagnoses of other nodes, data matrix, character list and other parameters.

Received 1 May; accepted 21 September 1998.

- Gregory, W. K. & Simpson, G. G. Cretaceous mammal skulls from Mongolia. *Am. Mus. Novit.* **225**, 7–20 (1926).
- Kielan-Jaworowska, Z. Evolution of the therian mammals in the Late Cretaceous of Asia. Part I. Deltatheriidae. *Palaeontol. Pol.* **33**, 103–131 (1975).
- Kielan-Jaworowska, Z. & Nessel, L. A. On the metatherian nature of the Deltatheroidea, a sister group of the Marsupalia. *Lethaia* **23**, 1–10 (1990).
- Averianov, A. O. New Late Cretaceous mammals of southern Kazakhstan. *Acta Palaeontol. Pol.* **42**, 243–256 (1997).
- Fox, R. *Deltatheroidea* like mammals from the Upper Cretaceous of North America. *Nature* **249**, 392 (1974).
- Cifelli, R. L. Cretaceous mammals of southern Utah. III. Therian mammals from the Turonian (early Late Cretaceous). *J. Vert. Paleontol.* **10**, 332–345 (1990).
- Van Valen, L. Deltatheriidae, a new order of mammals. *Bull. Am. Mus. Nat. Hist.* **132**, 1–126 (1966).
- McKenna, M. C., Mellett, J. S. & Szalay, F. S. Relationships of the Cretaceous mammal *Deltatheridium*. *J. Paleontol.* **45**, 441–442 (1971).
- Marshall, L. G. & Kielan-Jaworowska, Z. Relationships of the dog-like marsupials, deltatheroidans and early tribosphenic mammals. *Lethaia* **25**, 361–374 (1992).
- Szalay, F. S. *Evolutionary History of the Marsupials and an Analysis of Osteological Characters* (Cambridge Univ. Press, Cambridge, 1994).
- de Muizon, C., Cifelli, R. L. & Céspeles Paz, R. The origin of the dog-like borhyaenoid marsupials of South America. *Nature* **389**, 486–489 (1997).
- Butler, P. M. & Kielan-Jaworowska, Z. Is *Deltatheridium* a marsupial? *Nature* **245**, 105–106 (1973).
- Kielan-Jaworowska, Z., Eaton, J. G. & Bown, T. M. In *Mesozoic Mammals: the First Two Thirds of Mammalian History* (eds Lillegraven, J. A., Kielan-Jaworowska, Z. & Clemens, W. A.) 182–191 (Univ. Calif. Press, Berkeley, 1979).
- Cifelli, R. L. In *Mammal Phylogeny, Mesozoic Differentiation, Multituberculates Monotremes, Early Therians, and Marsupials* (eds Szalay, F. S., Novacek, M. J. & McKenna, M. C.) 205–215 (Springer, New York, 1993).

15. Cifelli, R. L. Early Cretaceous mammal from North America and the evolution of marsupial dental characters. *Proc. Natl Acad. Sci. USA* **90**, 9413–9416 (1993).
16. Dashzeveg, D. *et al.* Extraordinary preservation in a new vertebrate assemblage from the Late Cretaceous of Mongolia. *Nature* **374**, 446–449 (1995).
17. Hershkovitz, P. The staggered marsupial third lower incisor: hallmark of cohort Didelphimorphia, and description of a new genus and species with staggered I3 from the Albian (Lower Cretaceous) of Texas. *Boon. Zool. Beitrag* **45**, 153–169 (1995).
18. Fox, R. Molar structure and function in the Early Cretaceous mammal *Pappotherium*: evolutionary implications for Mesozoic Theria. *Can. J. Earth Sci.* **12**, 412–442 (1975).
19. Sanchez-Villagra, M. R. & Smith, K. K. Diversity and evolution of the marsupial mandibular angular process. *J. Mamm. Evol.* **4**, 119–144 (1997).
20. Wible, J. R. Petrosals of Late Cretaceous marsupials from North America, and a cladistic analysis of the petrosal in therian mammals. *J. Vert. Paleontol.* **10**, 183–205 (1990).
21. Luckett, W. P. in *Mammal Phylogeny, Mesozoic Differentiation, Multituberculates, Monotremes, Early Therians, and Marsupials* (eds Szalay, F. S., Novacek, M. J. & McKenna, M. C.) 182–204 (Springer, New York, 1993).
22. Cifelli, R. L. *et al.* Fossil evidence for the origin of the marsupial pattern of tooth replacement. *Nature* **379**, 715–718 (1996).
23. Cifelli, R. L. & de Muizon, C. Tooth eruption and replacement pattern in early marsupials. *C.R. Acad. Sci. Earth Planet. Sci.* **326**, 215–220 (1998).
24. Slaughter, B. H., Pine, R. H. & Pine, N. E. Eruption of cheek teeth in Insectivora and Carnivora. *J. Mamm.* **55**, 115–125 (1974).
25. Martin, T. Tooth replacement in Late Jurassic Dryolestidae (Eupantotheria, Mammalia). *J. Mamm. Evol.* **4**, 1–18 (1997).
26. Tyndale Biscoe, C. H. & Renfree, M. B. *Reproductive Physiology of Marsupials* (Cambridge Univ. Press, Cambridge, 1987).
27. Maier, W. in *Mammal Phylogeny, Mesozoic Differentiation, Multituberculates, Monotremes, Early Therians, and Marsupials* (eds Szalay, F. S., Novacek, M. J. & McKenna, M. C.) 165–181 (Springer, New York, 1993).
28. Kirsch, J. A. W., Laporte, E.-J. & Springer, M. S. DNA-hybridisation studies of marsupials and their implications for metatherian classification. *Aust. J. Zool.* **45**, 211–280 (1997).
29. Springer, M. S., Kirsch, J. A. W. & Case, J. A. in *Molecular Evolution and Adaptive Radiation* (eds Givnish, T. & Systma, K.) 129–161 (Cambridge Univ. Press, Cambridge, 1997).
30. Hedges, S. B., Parker, P. H., Sibley, G. & Kumar, S. Continental breakup and the ordinal diversification of birds and mammals. *Nature* **381**, 226–228 (1996).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. We thank A. Davidson for preparation of the fossils; L. Meeker, C. Tarka and E. Heck for illustrations; and J. Hopson, I. Horvitz, M. McKenna, C. de Muizon and M. Norell for comments on the paper. This work was supported by the NSF, the Jaffe Foundation, the National Geographic Society, the Epley Foundation, the Mercedes-Benz Corporation, the James Carter Memorial Fund and the Frick Laboratory Endowment of the American Museum of Natural History.

Correspondence and request for materials should be addressed to G.W.R. (e-mail: grougier@louisville.edu).

A role of *Ultrabithorax* in morphological differences between *Drosophila* species

David L. Stern*

Wellcome/CRC Institute, Tennis Court Road, Cambridge CB2 1QR, UK

The mechanisms underlying the evolution of morphology are poorly understood^{1,2}. Distantly related taxa sometimes exhibit correlations between morphological differences and patterns of gene expression^{3–8}, but such comparisons cannot establish how mechanisms evolve to generate diverse morphologies. Answers to these questions require resolution of the nature of developmental evolution within and between closely related species. Here I show how the detailed regulation of the Hox gene *Ultrabithorax* patterns trichomes on the posterior femur of the second leg in *Drosophila melanogaster*, and that evolution of *Ultrabithorax* has contributed to divergence of this feature among closely related species. The *cis*-regulatory regions of *Ultrabithorax*, and not the protein itself, appear to have evolved. This study provides experimental evidence that *cis*-regulatory evolution is one way in which conserved proteins have promoted morphological diversity¹.

In most species of the genus *Drosophila*, non-sensory microtrichia, or trichomes, cover much of the posterior second femur, leaving a patch of naked cuticle near the proximal end (Fig. 1). The distribution of this naked cuticle varies between, and to some extent

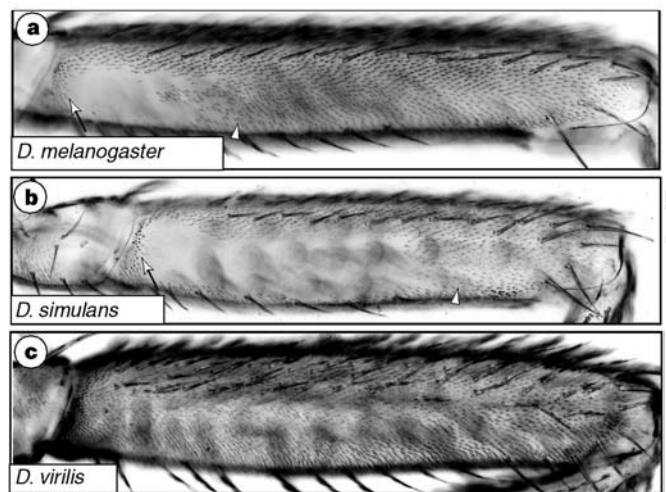


Figure 1 Trichome patterns on the posterior second femur vary among *Drosophila* species. **a**, *D. melanogaster* Oregon-R strain (mean naked cuticle length/femur length $\pm$ s.d., 0.31 ± 0.0035). **b**, *D. simulans* Tsimbazaza strain (0.61 ± 0.022). **c**, *D. virilis* Novosibirsk strain. The naked cuticle length was measured between the proximal (arrow) and maximum distal (arrowhead) extent of naked cuticle.

within, species. Of the three species studied here, *D. melanogaster* has a small naked patch, its sister species *D. simulans* has a larger patch, and the more distantly related *D. virilis* has no naked cuticle (Fig. 1).

In *D. melanogaster*, *Ultrabithorax* (*Ubx*) patterns unique morphological features from the second thoracic to the seventh abdomi-

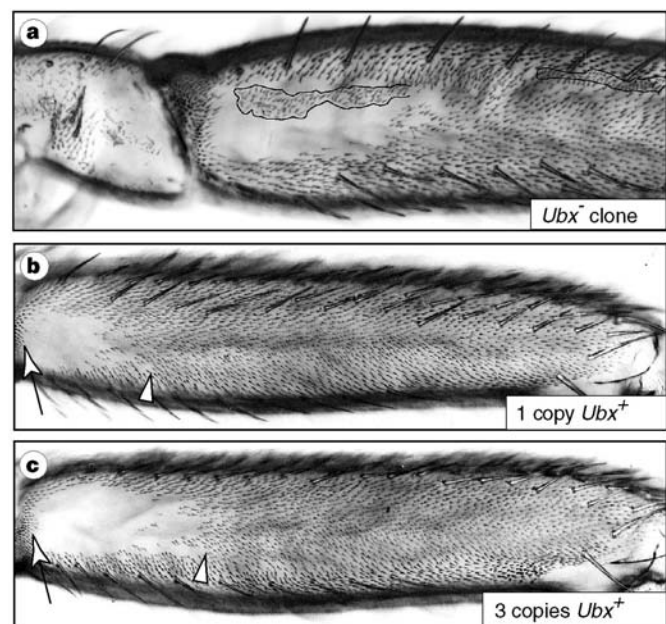


Figure 2 *Ubx* represses trichomes in the proximal naked cuticle in a dosage-dependent manner in *D. melanogaster*. **a**, A multiple-wing-hairs marked clone of *Ubx*⁻ cells (outlined) differentiated trichomes within the patch of naked cuticle. **b**, **c**, *Ubx* dosage altered the distribution of trichomes. Offspring from the cross *Df(3R)P9/Dp(3R)P5* $\times$ *stp*^{e11} with one functional copy (**b**) and three functional copies (**c**) of *Ubx* are shown. (Mean naked cuticle length/femur length $\pm$ s.d.: *Df(3R)P9/stp*^{e11} = 0.17 ± 0.014 versus *Dp(3R)P5/stp*^{e11} = 0.25 ± 0.014 ; $t = 7.98$, d.f. = 5, $P = 0.0005$.) Arrows and arrowheads delineate the extent of naked cuticle.

* Present address: Laboratory for Development and Evolution, University Museum of Zoology and Department of Zoology, Downing Street, Cambridge CB2 3EJ, UK.

nal segment^{9–13}. I tested the requirement of *Ubx* in patterning trichomes on the posterior second femur by generating clones of cells that lacked the ability to produce *Ubx* protein. When these clones were produced in the proximal patch of naked cuticle they differentiated trichomes; thus, *Ubx* is needed to repress trichomes in this region (Fig. 2a).

Three experiments indicate that the detailed expression pattern of *Ubx* is required to generate the specific morphology of a naked patch of cuticle. First, flies carrying three copies of the *Ubx* locus had

significantly more naked cuticle than sibling flies carrying one copy (Fig. 2b, c). Second, expression of uniform, high levels of *Ubx* protein during pupal development repressed trichomes on most of the posterior second femur (Fig. 3). Maximal repression occurred between 20 and 28 hours after puparium formation (APF). Trichomes were repressed in a proximal to distal direction, so that expression before 18 hours APF or at lower levels (results not shown) repressed proximal, but not distal, trichomes. Finally, *Ubx* protein is expressed in a proximal–distal gradient (Fig. 4a, b), with

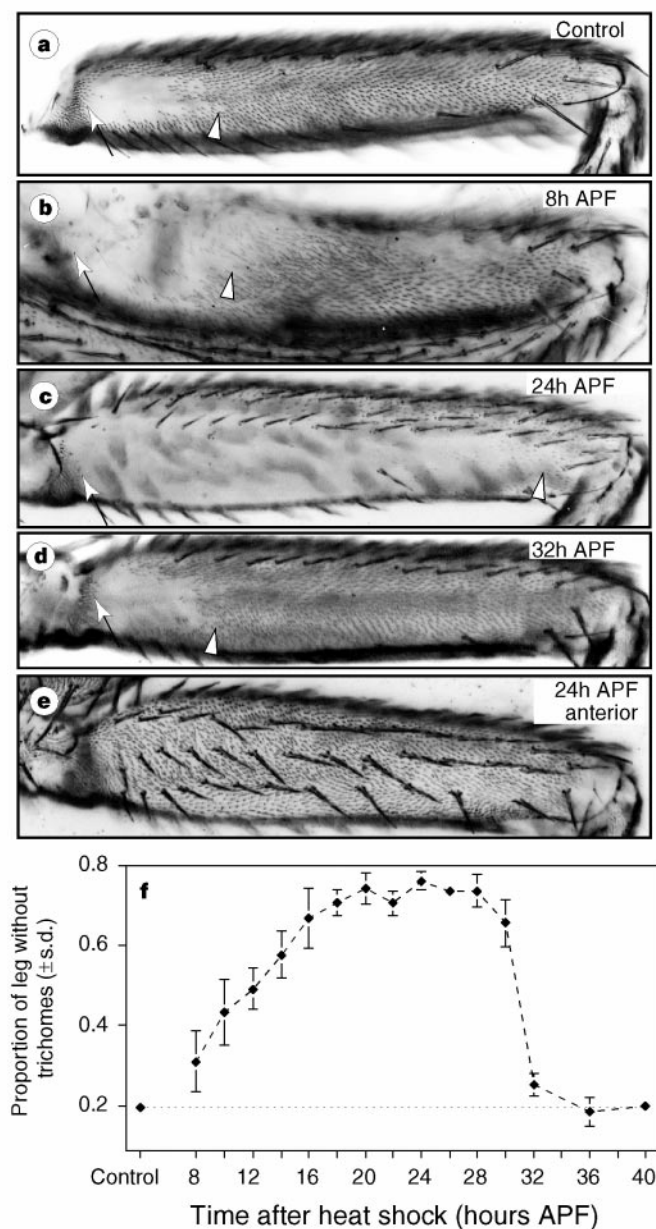


Figure 3 Uniform expression of *Ubx* in *D. melanogaster* represses trichomes on the posterior second femur during a short temporal window. **a**, Control flies. **b**, **c**, Flies heat-shocked at 8 h APF (**b**) and 24 h APF (**c**) showed progressively larger patches of naked cuticle. **d**, Heat shocks at 32 h APF had little effect. **e**, Ectopic *Ubx* failed to repress anterior femur trichomes. **f**, Ectopic *Ubx* repressed posterior second femur trichomes most efficiently between 18 and 28 h APF. (After heat shock, increased amounts of *Ubx* protein remained detectable for 7 h (results not shown). This long perdurance, combined with the sharp drop in sensitivity to *Ubx* after 30 h APF, indicates that cells may respond to *Ubx* level during a shorter time window than is seen here.) Arrows and arrowheads in **a–d** delineate the extent of naked cuticle.

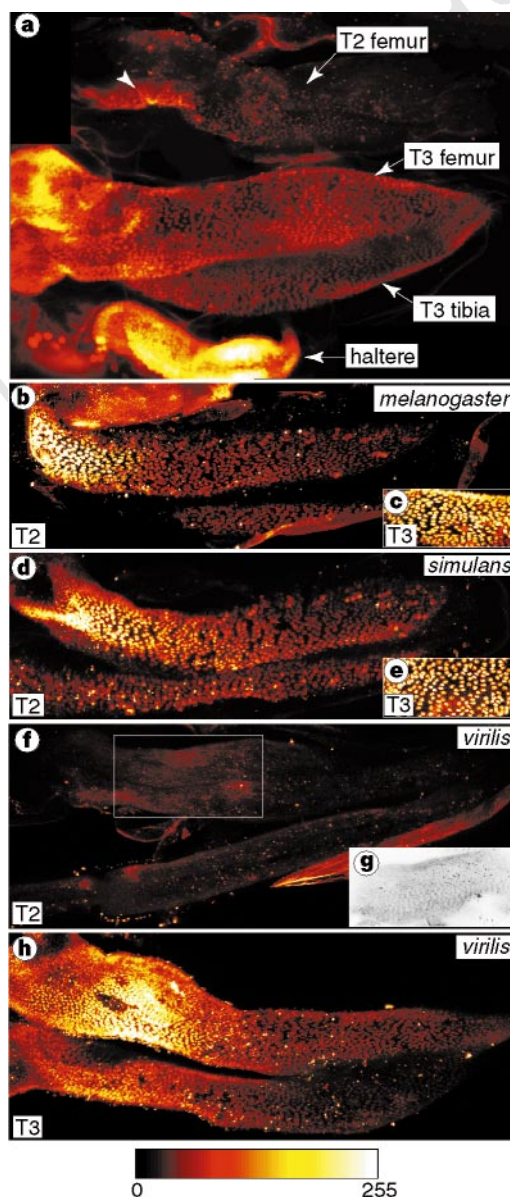


Figure 4 Distribution of *Ubx* protein in the posterior femurs of *Drosophila* species. **a**, Protein expression in the second (T2) and third (T3) leg and haltere in *D. melanogaster*. Proximal is to the left and the tibias fold back under the femurs. The second leg is twisted, providing an oblique view of the proximal femur (arrowhead) where *Ubx* is expressed at levels similar to those in the third femur. **b–e**, *D. melanogaster* (**b**) and *D. simulans* (**d**) show a proximal to distal *Ubx* expression gradient in T2. The approximate expression levels in the T3 femurs are shown for *D. melanogaster* (**c**) and *D. simulans* (**e**). **f–h**, *Ubx* expression in *D. virilis* is not visible in the second femur (**f**) at confocal settings that reveal high levels in the third femur (**h**). The boxed region of **f**, scanned at maximal sensitivity, shows that *Ubx* is expressed at low levels (**g**). Except in **g**, images were false-coloured with the scale shown below. Expression levels cannot be compared between **b**, **d** and **f**.

high levels proximally. Studies of the third leg support the hypothesis that high levels of *Ubx* repress trichomes. Clonal analysis showed that *Ubx* is required to repress trichomes on the posterior third leg (results not shown). In all three *Drosophila* species, most of the posterior third femur lacks trichomes and *Ubx* is expressed at high levels (Fig. 4a, c, e, h). The amount of expression in the posterior third femur is similar to that in the proximal patch of the posterior second femur, indicating that *Ubx* expression at or near these levels may repress posterior femur trichomes.

Patterns of *Ubx* expression correlate broadly with interspecific variation in trichome patterning. *Ubx* expression appeared to be similar in the two sister species *D. melanogaster* and *D. simulans* (Fig. 4b, d), both of which have a patch of naked cuticle. However, *D. virilis*, which has no naked cuticle (Fig. 1), expressed *Ubx* at levels far below those seen in the posterior third leg (Fig. 4f–h).

Existing methods of visualizing protein expression may be inadequate to detect the potentially small differences in the *Ubx* expression gradient between *D. melanogaster* and *D. simulans*. However, trichome production, which is a binary outcome of the amount of *Ubx* expression, may provide a more sensitive assay for differences in species' alleles. I therefore used a complementation

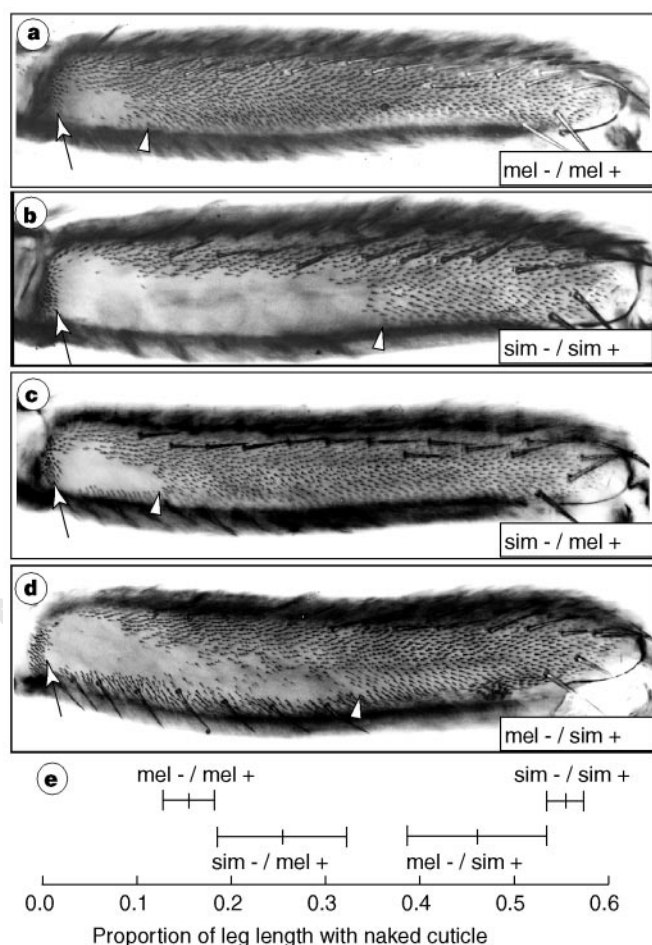


Figure 5 The *Ubx* alleles from *D. melanogaster* and *D. simulans* produce different trichome patterns in identical hybrid backgrounds. **a, b**, The posterior second femurs from *Ubx* heterozygous flies produced from parental strains of *D. melanogaster* (**a**) and *D. simulans* (**b**) in which new *Ubx* alleles had been induced. **c, d**, The posterior second femurs of the *Ubx* heterozygous F_1 offspring of the cross *sim - / sim +* $\times$ *mel + / mel +* (**c**) and *mel - / mel +* $\times$ *sim + / sim +* (**d**). **e**, The mean proportion of naked cuticle $\pm$ s.d. for the four genotypes. Arrows and arrowheads delineate the extent of naked cuticle. *mel -*, *D. melanogaster Ubx* null allele; *mel +*, *D. melanogaster Ubx* wild-type allele; *sim -*, *D. simulans Ubx* null allele; *sim +*, *D. simulans Ubx* wild-type allele.

Table 1 Nucleotide divergence between *D. melanogaster* and *D. simulans* for *Ubx* exons and flanking regions

	Coding		Flanking	
	Differences*	Total sites	Differences	Total sites
Exon 1	11	766	1	53
Exon 2	0	51	5	167
			(+ 1 insert/deletion)†	
Exon 3	0	51	1	53
Exon 4	5	302	6	96
Total	16 (1.4%)	1,170	13 (3.5%)	375
			(+ 1 insert/deletion)	

* All nucleotide differences in the coding regions were synonymous substitutions at third-base-pair positions.

† One 7-base-pair insertion/deletion was found downstream of exon 2.

test, in which a single functional *Ubx* allele from each species was tested in identical hybrid backgrounds. This approach bypassed the practical difficulties of analysis of F_2 generations of these species¹⁴.

To minimize the variation that arises from segregating modifier loci in existing stocks of *Ubx* mutants (results not shown), I generated new null alleles of *Ubx* in inbred lines of each species. Flies heterozygous for each new *Ubx* mutation were then crossed to the original inbred line of the other species (Fig. 5). F_1 offspring carrying a null *Ubx* allele from one species had a wild-type allele from the other and should be identical hybrids at all other loci. These F_1 classes differed significantly from each other in the predicted direction; flies carrying the *D. melanogaster* wild-type allele of *Ubx* (Fig. 5c) had a smaller naked patch than flies carrying the *D. simulans* wild-type allele (Fig. 5d, $t = 11.4$, degrees of freedom (d.f.) = 43, $P < 0.0001$). Therefore, the *Ubx* locus has evolved differently in these species so that the *D. simulans* allele produces a larger naked patch than the *D. melanogaster* allele. These conclusions are supported by complementation tests done using previously existing *Ubx* alleles (results not shown). The presumptive *Ubx* protein in *D. simulans* is identical to the *D. melanogaster* protein (Table 1), indicating that *cis*-regulation of the *Ubx* protein may have evolved¹⁵.

Ubx is not the only locus that influences this trait, because the hybrids were significantly different from the parental classes that carried the same wild-type allele of *Ubx* (Fig. 5; *D. melanogaster* null allele plus wild-type allele versus *D. simulans* null allele plus *D. melanogaster* wild-type allele: $t = -8.2$, d.f. = 58, $P < 0.0001$; *D. simulans* null plus wild-type alleles versus *D. melanogaster* null allele plus *D. simulans* wild-type allele: $t = -6.0$, d.f. = 25, $P < 0.0001$). The hybrids with a *D. simulans* wild-type allele also exhibited a dorsal–ventral pattern of trichome distribution (Fig. 5d): the ventral naked cuticle was longer than the dorsal cuticle. Therefore, at least one other gene, which is possibly involved in dorsal–ventral leg patterning, has contributed to the evolution of this trait.

This study illustrates one mechanism that contributes to a single difference between fly species. There are about 150,000 species of flies, all using roughly the same set of conserved genes to generate their morphology. How does such diversity arise from a conserved set of proteins? One possible answer, illustrated by my results, is that conserved proteins exhibit complex spatiotemporal regulation and that every element of this pattern is susceptible to subtle evolutionary manipulation^{1,6–8,15–17}.

Methods

Clonal analysis. Two clonal analysis¹⁸ experiments were performed. First, I crossed males carrying the alleles *mwh jv st red sbd² Ubx^{12.5} e¹¹ ro ca/TM1* to females carrying the alleles *f⁶⁴M(3)w f⁸⁷TM3*. Legs of male offspring without balancer chromosomes were studied for *f⁶⁴* bristles. Second, I crossed *mwh jv st red sbd² Ubx^{12.5} e¹¹ ro ca/TM1* males to *Dp(3:3)S462, Cbx¹ Ubx¹/TM6B* females. Legs of offspring without balancers were studied for *mwh* trichomes.

Larvae were subjected to X-rays (1,000 rad) at 24–72 h after egg-laying.

Ectopic *Ubx* expression. White prepupae from the cross *st^{p^e11}* × HS*Ubx* – 1a (ref. 19) were aged at 25 °C. Separate samples were heat-shocked (at 37 °C) for 1 h at 2-h intervals from 0–48 h APF. Samples heat-shocked at 0–6 h APF and some heat-shocked at 8 h APF failed to differentiate cuticle, whereas all others developed to pharate adults but failed to eclose. Most anterior femur trichomes were not repressed, indicating that neither ectopic *Ubx* expression nor the heat shock itself interfered with the ability of cells to differentiate trichomes.

Antibody staining. White prepupae were aged at 25 °C for 20–30 h for *D. melanogaster* and *D. simulans*, and for 38–44 h for *D. virilis*. Dissected pupal legs were stained using standard techniques²⁰, except that the antibodies were applied at 4 °C for at least 12 h each. The two monoclonal antibodies Fp3.38 (for *D. melanogaster* and *D. simulans*) and Fp6.87 (for *D. virilis*) were used together with a fluorescein isothiocyanate (FITC)-conjugated anti-mouse secondary antibody to detect *Ubx* protein^{20,21}. Both of the monoclonal antibodies produced comparable results when tested in *D. melanogaster* (results not shown). Images were captured with a BioRad MRC-1024 confocal microscope.

Generation of new *Ubx* alleles. New alleles of *Ubx* were induced in the *st^{p^e11}* *D. melanogaster* and Tsimbazaza *D. simulans* stocks by X-ray irradiation of 3-day-old virgin males with 4 krad. These males were mated with virgin 3-day-old females. Offspring were screened for enlarged halteres. One of 3,600 *D. melanogaster* and nine of 8,500 *D. simulans* flies had enlarged halteres, of which one and two flies of each species, respectively, were fertile. Only alleles that failed to complement *Ubx*, showed the appropriate null phenotype⁹, and did not produce *Ubx* protein (as assessed by antibody staining) were used for the interspecific test.

DNA sequencing. The four *Ubx* exons were amplified by the polymerase chain reaction (PCR) from genomic DNA extracted from the Tsimbazaza strain of *D. simulans*, with the following primers: exon 1, 5'-GCCCGTCTCAGACGG AGCAC-3' and 5'-TGGGATTTCGGGGGACTTTCAG-3'; exon 2, 5'-CCCTA CCACAGATCCCCACGTACCC-3' and 5'-GCCCATTCATTCATGAATTT AGCACACC-3'; exon 3, 5'-TGAGGCATATGACGTTCTCTGGAC-3' and 5'-CATAGGCAAACTATAGGCTAAGGGTTTAC-3'; exon 4, 5'-ATGTATGTA TTTCTGTCGATGCAGGTC-3' and 5'-CCAATCCCACATACACCCTAC-3'. PCR fragments were cloned into pCRII (Invitrogen) and four clones of each exon were sequenced using Dye Terminator Cycle sequencing (Perkin Elmer).

Received 10 August; accepted 14 October 1998.

- Carroll, S. B. Homeotic genes and the evolution of arthropods and chordates. *Nature* **376**, 479–485 (1995).
- Gerhart, J. & Kirschner, M. *Cells, Embryos and Evolution* (Blackwell, Malden, Massachusetts, 1997).
- Warren, R. W., Nagy, L., Selegue, J., Gates, J. & Carroll, S. Evolution of homeotic gene regulation and function in flies and butterflies. *Nature* **372**, 458–461 (1994).
- Akam, M. Hox genes and the evolution of diverse body plans. *Phil. Trans. R. Soc. Lond. B* **349**, 313–319 (1995).
- Averof, M. & Akam, M. Hox genes and the diversification of insect and crustacean body plans. *Nature* **376**, 420–423 (1995).
- Averof, M. & Patel, N. H. Crustacean appendage evolution associated with changes in Hox gene expression. *Nature* **388**, 682–686 (1997).
- Grenier, J. K., Garber, T. L., Warren, R., Whittington, P. M. & Carroll, S. Evolution of the entire arthropod Hox gene set predated the origin and radiation of the onychophoran/arthropod clade. *Curr. Biol.* **7**, 547–553 (1997).
- Carroll, S. J. & Wray, G. A. Radical alterations in the roles of homeobox genes during echinoderm evolution. *Nature* **389**, 718–721 (1997).
- Lewis, E. B. A gene complex controlling segmentation in *Drosophila*. *Nature* **276**, 565–570 (1978).
- Morata, G. & Kerridge, S. Sequential functions of the bithorax complex of *Drosophila*. *Nature* **290**, 778–781 (1981).
- Sanchez-Herrero, E., Casanova, J., Kerridge, S. & Morata, G. Anatomy and function of the bithorax complex of *Drosophila*. *Cold Spring Harb. Symp. Quant. Biol.* **50**, 165–172 (1985).
- Gonzalez-Reyes, A. & Morata, G. The developmental effect of overexpressing a *Ubx* product in *Drosophila* embryos is dependent on its interactions with other homeotic products. *Cell* **61**, 515–522 (1990).
- Castelli-Gair, J. & Akam, M. How the Hox gene *Ultrabithorax* specifies two different segments: the significance of spatial and temporal regulation within metameres. *Development* **121**, 2973–2982 (1995).
- Davis, A. W. *et al.* Rescue of hybrid sterility in crosses between *D. melanogaster* and *D. simulans*. *Nature* **380**, 157–159 (1996).
- Gibson, G. & Hogness, D. S. Effect of polymorphism in the *Drosophila* regulatory gene *Ultrabithorax* on homeotic stability. *Science* **271**, 200–203 (1996).
- Akam, M. Hox genes, homeosis and the evolution of segment identity; no need for hopeless monsters. *Int. J. Dev. Biol.* **42**, 445–451 (1998).
- Castelli-Gair, J. Implications of the spatial and temporal regulation of Hox genes on development and evolution. *Int. J. Dev. Biol.* **42**, 437–444 (1998).
- Morata, G. & Ripoll, P. Minutes: mutants of *Drosophila* autonomously affecting cell division rate. *Dev. Biol.* **42**, 211–221 (1975).
- Mann, R. S. & Hogness, D. S. Functional dissection of *Ultrabithorax* proteins in *D. melanogaster*. *Cell* **60**, 597–610 (1990).

- Kelsh, R., Weinzierl, R. O. J., White, R. A. H. & Akam, M. Homeotic gene expression in the locust *Schistocerca*: an antibody that detects conserved epitopes in *Ultrabithorax* and *Abdominal-A* proteins. *Dev. Genet.* **15**, 19–31 (1994).
- White, R. A. H. & Wilcox, M. Protein products of the bithorax complex in *Drosophila*. *Cell* **39**, 163–171 (1984).

Acknowledgements. I thank M. Akam for his patronage; N. Brown, J. Castelli-Gair and J. De Celis for fly tutelage; J. De Celis, E. Lewis, J. Roote and M. Ashburner for stocks; B. Yen and E. Sucena for technical assistance; and M. Akam, M. Ashburner, A. Gonzalez-Reyes, C. Mirth, L. Partridge and E. Sucena for comments. This work was supported by Churchill College, the Wellcome Trust and the BBSRC.

Correspondence and requests for materials should be addressed to the author. The *D. simulans* *Ubx* sequences have been deposited in Genbank (accession numbers AF099980–AF099983).

HMG-CoA reductase guides migrating primordial germ cells

Mark Van Doren, Heather Tarczy Broihier, Lisa A. Moore & Ruth Lehmann

Developmental Genetics Program, Skirball Institute and Howard Hughes Medical Institute, Department of Cell Biology, New York University Medical Center, 540 First Avenue, New York, New York 10016, USA

The enzyme 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase is best known for catalysing a rate-limiting step in cholesterol biosynthesis, but it also participates in the production of a wide variety of other compounds¹. Some clinical benefits attributed to inhibitors of HMG-CoA reductase are now thought to be independent of any serum cholesterol-lowering effect^{2,3}. Here we describe a new cholesterol-independent role for HMG-CoA reductase, in regulating a developmental process: primordial germ cell migration. We show that in *Drosophila* this enzyme is highly expressed in the somatic gonad and that it is necessary for primordial germ cells to migrate to this tissue. Misexpression of HMG-CoA reductase is sufficient to attract primordial germ cells to tissues other than the gonadal mesoderm. We conclude that the regulated expression of HMG-CoA reductase has a critical developmental function in providing spatial information to guide migrating primordial germ cells.

In many animals, including mammals and *Drosophila*, primordial germ cells (PGCs) form in a region of the embryo separate from the somatic portion of the gonad. Consequently, PGCs migrate through the embryo and make specific contacts with somatic cells to form the gonad^{4,5}. In a genetic screen for mutations that disrupt this process in *Drosophila*, we identified 15 alleles of a gene that we called *columbus* (*clb*)⁶. We have cloned *clb* and find that it encodes a *Drosophila* homologue of HMG-CoA reductase⁷ (*hmgr*). Twelve ethyl methane sulphonate (EMS)-induced alleles and one transposon-induced allele of *clb* were examined and all show molecular defects in the *hmgr* transcript (Fig. 1, and data not shown): we

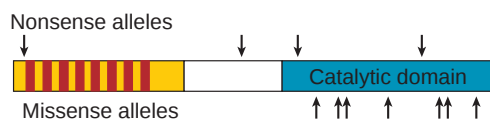


Figure 1 *columbus* is *Drosophila* HMG-CoA reductase. The HMG-CoA reductase open reading frame is shown schematically. The conserved amino terminus (yellow with transmembrane domains in red) and the conserved carboxy-terminal catalytic domain (blue) are 45% and 56% identical to human HMG-CoA reductase, respectively. Mutations found in 11 EMS-induced alleles are indicated by arrows (nonsense alleles above and missense alleles below the open reading frame). One EMS-induced allele has mutations in both a conserved amino acid in the catalytic domain and the transmembrane domain and is not represented in the figure.

therefore refer to our *clb* alleles as *hmgcr^{clb}*. All missense mutations found in *hmgcr* alter amino acids in the catalytic domain that are conserved from humans to archaeobacteria (Fig. 1), suggesting that it is the catalytic activity of HMG-CoA reductase that is important for PGC migration.

Mutations in *hmgcr* severely disrupt primordial germ cell migration (Fig. 2). In wild-type *Drosophila* embryos, PGCs first migrate through the posterior endoderm, move to the dorsal surface of the endoderm, and then migrate into the adjacent mesoderm (Fig. 2b; for review, see ref. 5). Once in the mesoderm, the PGCs associate with a subset of mesodermal cells that will give rise to the somatic portion of the gonad (gonadal mesoderm; Fig. 2d, f). Subsequently, the PGCs and gonadal mesoderm coalesce to form the embryonic gonad. In *hmgcr^{clb}* mutants, some PGCs fail to migrate from the endoderm to the mesoderm, and instead remain associated with the endoderm (Fig. 2a). In addition, many PGCs that do migrate to the mesoderm fail to associate with the gonadal mesoderm (Fig. 2e), and instead scatter widely in the embryo (Fig. 2c). Thus, HMG-CoA reductase is required for PGCs to migrate to the mesoderm and to find their target tissue within the mesoderm.

Several mutants that show defects in PGC migration affect development of the gonadal mesoderm^{6,8}. We therefore examined gonad formation in *hmgcr^{clb}* mutant embryos. Early specification of

gonadal mesoderm precursors occurs normally in *hmgcr^{clb}* mutants (data not shown), as judged by the expression of Zinc-finger homeodomain protein-1 (ref. 8) and Clift⁹. At later stages, *hmgcr^{clb}* mutant embryos continue to express gonadal mesoderm-specific markers such as Clift (Fig. 2e), *DWnt-2* (ref. 10) (Fig. 2g), and the 412 retrotransposon¹¹ (data not shown). In addition, the gonadal mesoderm coalesces normally into a correctly patterned embryonic gonad (Fig. 2g) which contains few or no PGCs. Thus, in contrast to previously described mutants, the PGC migration defect observed in *hmgcr^{clb}* mutant embryos cannot be attributed to a failure in gonadal mesoderm development. We have also found that the development of several other tissues occurs normally in *hmgcr^{clb}* mutant embryos (data not shown), including the nervous system and tracheal branches, which require directed cell movement for their formation. It is still possible that these processes are influenced by maternal HMG-CoA reductase which is required for early embryonic development¹². Although *hmgcr^{clb}* mutant embryos do not have any gross morphological defects, they do die at the end of embryogenesis. This is consistent with HMG-CoA reductase being required for some more general cellular function(s), in addition to being specifically required for PGC migration.

hmgcr is expressed in the embryo in a dynamic and tissue-specific pattern (Fig. 3). Although *hmgcr* is not expressed in PGCs (Fig. 3a), it is expressed in their target tissue, the gonadal mesoderm. In the mesoderm, *hmgcr* is initially broadly expressed (Fig. 3a) and then becomes restricted to a segmental pattern (Fig. 3b); this coincides with the stage when PGCs migrate from the endoderm into the mesoderm. Mesodermal *hmgcr* expression is then further restricted to a cluster of cells in each of parasegments 10, 11 and 12 (Fig. 3c),

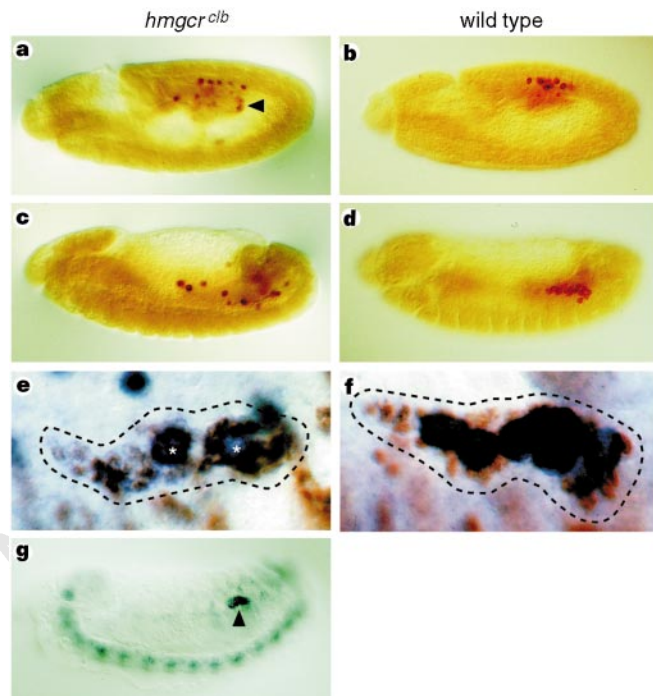


Figure 2 The phenotype of *hmgcr^{clb}* mutant embryos. **a-d**, Embryos immunostained with anti-Vasa, which labels the PGCs. All embryos have dorsal up, anterior left. Embryos shown are *hmgcr^{clb}* mutant stage 11 (**a**) and stage 13 (**c**), and wild-type stage 11 (**b**) and stage 13 (**d**). The arrowhead in **a** indicates PGCs that remain associated with the endoderm rather than migrating into the mesoderm. On average, 25% (range 0–40%, $n = 10$) of all PGCs in *hmgcr^{clb}* mutant embryos are associated with the endoderm at stage 13 compared with 2% (0–6%, $n = 10$) in wild type. **e, f**, High-magnification view of *hmgcr^{clb}* mutant (**e**) and wild-type (**f**) embryos immunostained with anti-Vasa (purple) and gonadal mesoderm-specific anti-Clift (brown). The gonadal mesoderm is outlined by the dashed line. Asterisks in **e** indicate PGCs that are associated with Clift-expressing cells. An average of 3.3 ± 1.2 PGCs associate with Clift-expressing cells in *hmgcr^{clb}* mutant embryos compared with 7.2 ± 1.9 in wild type. **g**, An *hmgcr^{clb}* mutant embryo labelled by *in situ* hybridization to *DWnt-2* RNA. *DWnt-2* expression is seen in the gonad (arrowhead) and is highest at the posterior (right); it is indistinguishable from wild-type expression. Staging according to ref. 20.

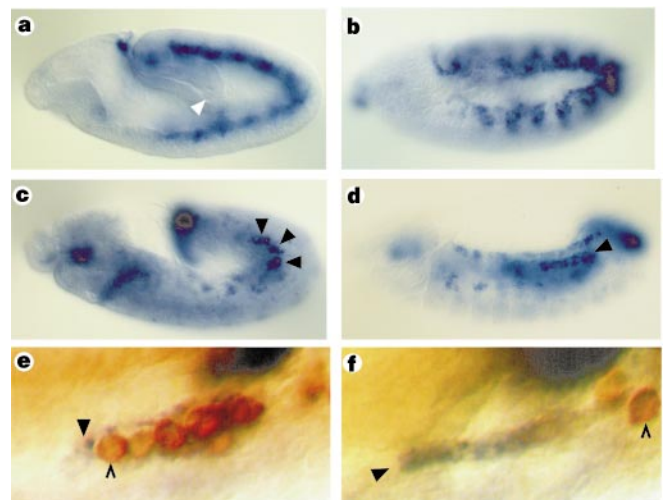


Figure 3 *hmgcr* is highly expressed in gonadal mesoderm. **a-d**, Wild-type embryos labelled by *in situ* hybridization to the *hmgcr* transcript orientated with dorsal up, anterior left. **a**, Stage 9. The mesoderm expresses *hmgcr* but PGCs (contained within the posterior midgut pocket; arrowhead) do not. **b**, Stage 11. Mesodermal expression is restricted to a segmental pattern. **c**, Stage 12. Arrowheads indicate the positions of the developing gonadal mesoderm clusters in parasegments 10, 11 and 12. **d**, Stage 13. Arrowhead indicates the gonadal mesoderm. *hmgcr* expression is also observed in the precursors of the dorsal vessel (line of cells dorsal to the gonadal mesoderm in **d**), in regions of the foregut, midgut and hindgut, in clusters of cells along the ventral midline, and in the salivary gland. Morphogenesis of the gut and salivary gland in *hmgcr* mutants is normal and the larval cuticle shows no patterning defect. **e**, Wild-type, and **f**, *hmgcr^{clb}* mutant embryos doubly labelled by *in situ* hybridization to the *hmgcr* transcript (purple) and by immunostaining with anti-Vasa (brown). The PGCs (examples indicated by open arrowheads) are closely aligned with *hmgcr*-expressing cells (closed arrowheads) in wild-type embryos, but not in *hmgcr^{clb}* mutants that express an RNA containing a nonsense mutation.

corresponding to where the developing gonadal mesoderm is found^{8,9}. During this time, the PGCs complete their association with the gonadal mesoderm. The three clusters of *hmgr*-expressing cells join to form a continuous band of cells (Fig. 3d–f), which is tightly associated with the PGCs in wild-type embryos (Fig. 3e), but not in *hmgr^{clb}* mutant embryos (Fig. 3f). Thus, *hmgr* is expressed at high levels in the developing gonadal mesoderm (compare Figs 3e and 2f) at the time the PGCs specifically associate with this tissue.

As *hmgr* is specifically expressed in the gonadal mesoderm and is required for PGCs to associate with this tissue properly, we investigated whether *hmgr* guides PGCs to the gonadal mesoderm. This was done by ectopically expressing *hmgr* in specific tissues using the Gal4/UAS system¹³. When *hmgr* is expressed in stripes within the mesoderm and ectoderm, many PGCs become lost and are preferentially found in places of highest ectopic *hmgr* expression (Fig. 4a, b). These embryos are viable and the gonadal mesoderm develops normally (data not shown). Thus the correct pattern of *hmgr* expression within the mesoderm is essential for the PGCs to associate

with the gonadal mesoderm. To test whether *hmgr* expression is sufficient to attract PGCs to new locations, we expressed *hmgr* in the epidermis (Fig. 4c) and in the nervous system (Fig. 4d). In both cases, PGCs now associated with the tissue ectopically expressing *hmgr*. Therefore, ectopic expression of *hmgr* is sufficient to attract germ cells to tissues to which they would never normally migrate. In contrast, when we expressed *hmgr* in PGCs, their migration was unaffected (Fig. 4e) and the embryos grew up to be fertile adults. We conclude that high-level expression of *hmgr* in the gonadal mesoderm provides spatial information to guide migrating PGCs.

Our results indicate that *hmgr* is limiting for production of a signal emanating from the gonadal mesoderm to attract migrating PGCs. Although some products of the HMG-CoA reductase pathway should be essential in all cells, increased expression of this enzyme in the gonadal mesoderm is necessary for signalling to PGCs. The gene *wunen* encodes a repulsive factor for migrating PGCs¹⁴; thus, a combination of positive and negative cues controls PGC migration in *Drosophila*. It will be interesting to see whether factors such as HMG-CoA reductase and Wunen play a role in PGC migration in other systems. An important question is the nature of the product downstream of HMG-CoA reductase that is responsible for the attractive signal: it cannot be cholesterol or one of its derivatives as *Drosophila* does not synthesize sterols by this pathway¹⁵. Other candidates include a wide variety of terpene derivatives, as well as the products necessary for *N*-glycosylation and isoprenylation of proteins. Products of phospholipid biosynthesis act as extracellular signals and chemoattractants^{16,17}, and the products of the HMG-CoA reductase pathway may act similarly. Indeed, inhibitors of HMG-CoA reductase can interfere with mammalian cell migration and chemotaxis by acting on both the migrating cells² and the target tissue¹⁸. Our results provide evidence that the HMG-CoA reductase pathway can be used to guide cell migration and offer a new context for the action of HMG-CoA reductase in development and disease. □

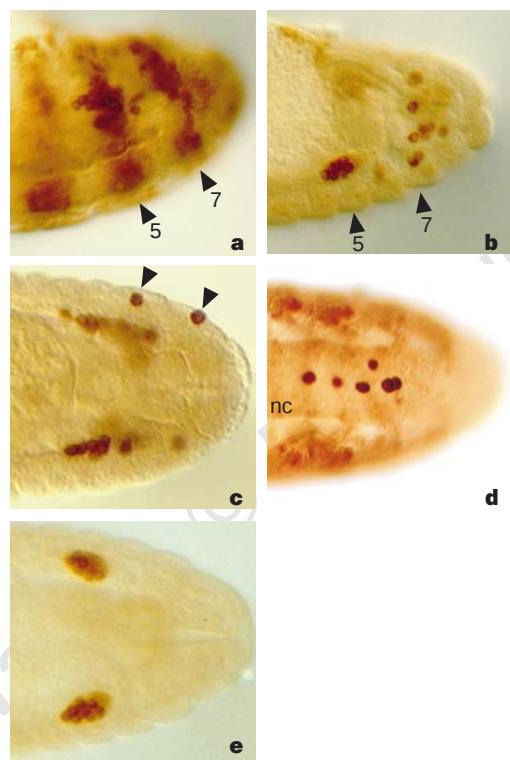


Figure 4 Ectopic *hmgr* expression is sufficient to attract PGCs to new locations. All embryos are immunostained with anti-Vasa and contain *UAS-lacZ* (a) or *UAS-hmgr* (b–e) with tissue-specific Gal4 expression, as indicated. **a, b**, Lateral view; **c–e**, frontal view. **a**, Embryo containing *hairy-Gal4* and *UAS-lacZ* immunostained with anti-β-galactosidase. The segmental staining is indicated by arrowheads at segments A5 and A7. **b**, Embryo expressing *hmgr* in the pattern seen in **a**. Although some PGCs are found in the gonad, many are lost and accumulate preferentially in segment A7 (7.4 ± 3.1 PGCs lost in A7, 2.9 ± 1.8 in A6, 1.8 ± 1.7 in A5; $n = 15$ embryos). PGCs are more rarely lost in A5, presumably because they are attracted to the gonadal mesoderm present in this segment and so are incorporated into the gonad. As PGCs initially enter posterior mesoderm, lost PGCs are rarely seen anterior to A5. **c**, Embryo expressing *hmgr* in the epidermis (*69B-Gal4*). PGCs are found in the outermost tissue layer of the embryo (arrowheads). **d**, Embryo expressing *hmgr* in the nervous system (*elav-Gal4*). PGCs associate with the ventral nerve cord (nc). **e**, Embryo expressing *hmgr* in the PGCs (*nos-Gal4VP16*). PGC migration is normal. Embryos containing *UAS-hmgr* and *hairy-Gal4*, *elav-Gal4* or *nos-Gal4VP16* can all grow into viable adults (*69B-Gal4* was not tested). Adults containing *UAS-hmgr* with *elav-Gal4* or *nos-Gal4VP16* are also fertile (others were not tested).

Methods

Cloning *clb* and identification as *hmgr*. *clb* was meiotically mapped to position 3–80 (ref. 6) and failed to complement the transposon-induced mutation *l(3)01152⁰¹¹⁵²*. Excision of this transposon reverts the lethality *in trans* to *clb*. Genomic DNA next to the *l(3)01152⁰¹¹⁵²* transposon was isolated and additional genomic DNA was obtained from the Berkeley *Drosophila* Genome Project. This DNA was used to probe northern blots and a cDNA library (a gift from N. Brown). The *l(3)01152⁰¹¹⁵²* transposon is inserted 179 nucleotides into the 5' end of the *hmgr* transcript. EMS-induced *clb* alleles were sequenced after PCR amplification of genomic DNA from homozygous mutant embryos.

Whole-mount *in situ* hybridization and immunocytochemistry. These procedures have been described⁸. Antisense probes were generated by digesting our cDNA pNB36C with *Hind*III and transcribing with T7 RNA polymerase (*hmgr*), and by digesting plasmid pBSDWnt-2 (a gift from R. Nusse) with *Bam*HI and transcribing with T3 RNA polymerase (*DWnt-2*). Antibodies used were anti-Vasa (1:5,000; a gift from A. Williamson), anti-β-galactosidase (1:20,000; Capel) and anti-Clift (1:1,000; a gift from N. Bonini).

Mutant and transgenic *Drosophila* stocks. *hmgr* mutants used for phenotypic analysis are either *hmgr^{db1}* (an early nonsense mutation) homozygous or trans-heterozygous with *hmgr^{clb2}* (a strong missense mutation). Similar results are observed with either genotype. The *l(3)01152⁰¹¹⁵²* mutant was obtained from the Bloomington Stock Center. To generate pUAS-*hmgr*, the pNB36C *hmgr* cDNA was digested with *Pvu*II and blunt-end-ligated into the *Eco*RI site of the vector pUAST¹³. The resulting plasmid, pUAS-*hmgr*, was introduced into the *Drosophila* genome using standard P-element-mediated transformation techniques. Flies expressing *lacZ* under Gal4 control (*UAS-lacZ*) were Bg4-1–2 (ref. 13). The Gal4 lines used were: *hairy-Gal4* (IJ3, segmental¹³), *elav-Gal4* (nervous system; a gift from B. Jones) and *nos-Gal4VP16* (PGCs¹⁹).

Received 17 August; accepted 18 September 1998.

- Goldstein, J. L. & Brown, M. S. Regulation of the mevalonate pathway. *Nature* **343**, 425–430 (1990).
- Corsini, A. et al. Non-lipid-related effects of 3-hydroxy-3-methylglutaryl coenzyme A reductase inhibitors. *Cardiology* **87**, 458–468 (1996).

3. Katznelson, S. & Kobashigawa, J. A. Dual roles of HMG-CoA reductase inhibitors in solid organ transplantation: lipid lowering and immunosuppression. *Kidney Int. (suppl.)* **52**, 112–115 (1995).
4. Gomperts, M., Wylie, C. & Heasman, J. Primordial germ cell migration. *Ciba Foundn Symp. (Netherlands)* **182**, 121–134 (1994).
5. Williamson, A. & Lehmann, R. Germ cell development in *Drosophila*. *Annu. Rev. Cell Dev. Biol.* **12**, 365–391 (1996).
6. Moore, L. A., Broihier, H. T., Van Doren, M., Lunsford, L. B. & Lehmann, R. Identification of genes controlling germ cell migration and embryonic gonad formation in *Drosophila*. *Development* **125**, 667–678 (1998).
7. Gertler, F. B., Chiu, C. Y., Richter-Mann, L. & Chin, D. J. Developmental and metabolic regulation of the *Drosophila melanogaster* 3-hydroxy-3-methylglutaryl coenzyme A reductase. *Mol. Cell. Biol.* **8**, 2713–2721 (1988).
8. Broihier, H. T., Moore, L. A., Van Doren, M., Newman, S. & Lehmann, R. *zfh-1* is required for germ cell migration and gonadal mesoderm development in *Drosophila*. *Development* **125**, 655–666 (1998).
9. Boyle, M., Bonini, N. & DiNardo, S. Expression and function of clift in the development of somatic gonadal precursors within the *Drosophila* mesoderm. *Development* **124**, 971–982 (1997).
10. Russell, J., Gennissen, A. & Nusse, R. Isolation and expression of two novel Wnt/wingless gene homologues in *Drosophila*. *Development* **115**, 475–485 (1994).
11. Brookman, J., Toosy, A., Shashidhara, L. & White, R. The 412 retrotransposon and the development of gonadal mesoderm in *Drosophila*. *Development* **116**, 1185–1192 (1992).
12. Perrimon, N., Lanjuin, A., Arnold, C. & Noll, E. Zygotic lethal mutations with maternal effect phenotypes in *Drosophila melanogaster*. II. Loci on the second and third chromosomes identified by P-element-induced mutations. *Genetics* **144**, 1681–1692 (1996).
13. Brand, A. H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401–415 (1993).
14. Zhang, N., Zhang, J., Purcell, K. J., Cheng, Y. & Howard, K. The *Drosophila* protein Wunen repels migrating germ cells. *Nature* **385**, 64–67 (1997).
15. Clayton, R. B. The utilization of sterols by insects. *J. Lipid Res.* **5**, 3–19 (1964).
16. Cunningham, F. M. *Lipid Mediators* (Academic, San Diego, 1994).
17. Moolenaar, W. H. LPA: a novel lipid mediator with diverse biological actions. *Trends Cell Biol.* **4**, 213–219 (1994).
18. Kim, S. Y. *et al.* Human mesangial cell production of monocyte chemoattractant protein-1: modulation by lovastatin. *Kidney Int.* **48**, 363–371 (1995).
19. Van Doren, M., Williamson, A. & Lehmann, R. Regulation of zygotic gene expression in *Drosophila* primordial germ cells. *Curr. Biol.* **8**, 243–246 (1998).
20. Campos-Ortega, J. A. & Hartenstein, V. *The Embryonic Development of Drosophila melanogaster* (Springer, Berlin, 1997).

Acknowledgements. We thank D. Andrew, N. Bonini, N. Brown, B. Jones, R. Nusse, N. Perrimon, A. Williamson, the Bloomington Stock Center and the Berkeley *Drosophila* Genome Project for reagents; N. Perrimon, M. Pettitt and S. Abmayr for sharing unpublished data; R. Brewster, S. Burden, G. Fishell, D. Littman, E. Skolnik and members of our laboratory for critically reading the manuscript; and M. Samuels for technical assistance. M.V.D. was supported by the American Cancer Society and the Howard Hughes Medical Institute; R.L. is an investigator of the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to R.L. (e-mail: lehmann@saturn.med.nyu.edu).

Gain control of NMDA-receptor currents by intracellular sodium

Xian-Min Yu^{†‡§} & Michael W. Salter^{*†}

^{*} Programme in Brain and Behaviour, Hospital for Sick Children, 555 University Avenue, Toronto, Ontario M5G 1X8, Canada

Departments of [†] Physiology and [‡] Oral Physiology, University of Toronto, Toronto, Ontario, Ontario M5G 1G6, Canada

[§] Molecular Neurobiology Section, Clarke Institute of Psychiatry, Toronto, Ontario, Ontario M5T 1R8, Canada

The influx of Na⁺ is fundamental to electrical signalling in the nervous system and is essential for such basic signals as action potentials and excitatory postsynaptic potentials¹. During periods of bursting or high levels of discharge activity, large increases in intracellular Na⁺ concentration ([Na⁺]_i) are produced in neuronal soma and dendrites^{2–4}. However, the intracellular signalling function of raised postsynaptic [Na⁺]_i is unknown. Here we show that [Na⁺]_i regulates the function of NMDA (N-methyl-D-aspartate) receptors, a principal subtype of glutamate receptor⁵. NMDA-receptor-mediated whole-cell currents and NMDA-receptor single-channel activity were increased by raising [Na⁺]_i and channel activity decreased upon lowering [Na⁺]_i; therefore, the activity of NMDA channels tracks changes in [Na⁺]_i. We found that the sensitivity of the channel to Na⁺ was set by a Src kinase that is associated with the channel. Raising [Na⁺]_i selectively increased synaptic responses mediated by NMDA receptors, but

not by non-NMDA receptors. Thus, the change in postsynaptic [Na⁺]_i that occurs during neuronal activity is a signal for controlling the gain of excitatory synaptic transmission. This mechanism may be important for NMDA-receptor-dependent plasticity and toxicity in the central nervous system.

We used the whole-cell patch-clamp method to record currents from cultured mammalian spinal or hippocampal neurons. NMDA-receptor-mediated responses were evoked by brief applications of NMDA using pressure from a micropipette located near the neuron. We found that intracellular perfusion through the patch pipette with solution containing 50 mM Na⁺ increased the amplitude of NMDA currents (Fig. 1a, b). Ratiometric measurement of [Na⁺]_i, made using the Na⁺-sensitive dye sodium-binding benzofuran isophthalate (SBFI), indicated that the baseline level was 9.0 ± 2 mM and that the intracellular Na⁺ perfusion raised [Na⁺]_i only by 30 ± 3 mM (*n* = 4 cells), presumably because of the high activity of the Na⁺/K⁺ ATPase in neurons³. The slope of the current–voltage relationship for the NMDA currents increased during intracellular application of Na⁺, with no change in the reversal potential (Fig. 1c). Thus, the increase in the amplitude of NMDA currents was due to an increase in whole-cell conductance rather than to a change in driving force. In contrast, applying 50 mM Cs⁺ intracellularly did not affect the amplitude (Fig. 1b) or the reversal potential of the NMDA currents.

To characterize the effects of [Na⁺]_i on NMDA single-channel currents, we made cell-attached recordings while varying [Na⁺]_i by using the Na⁺ ionophore monensin^{6,7}. When the extracellular Na⁺ concentration ([Na⁺]_o) was 50 mM, application of monensin raised [Na⁺]_i by 24 ± 2 mM (*n* = 6 cells) and significantly increased the NMDA single-channel overall open probability (*P*_o) and mean open time (*t*_o) (Fig. 1d, e, g). The lengths of bursts and clusters were increased to 143 ± 16% and 178 ± 32% of control, respectively. However, there was no change in the single-channel conductance or the reversal potential (Fig. 1f). To confirm that the increase in channel activity during application of monensin was due to activity of NMDA channels, we administered the NMDA-receptor blocker MK801 (ref. 8), which abolished the single-channel currents (Fig. 1d, e). When [Na⁺]_i was raised by 7 ± 2 mM (*n* = 4 cells) there was a smaller increase in *P*_o and *t*_o (Fig. 1g), and when [Na⁺]_i was reduced by applying monensin in Na⁺-free bath solution, NMDA-channel activity decreased (Fig. 1g). Thus, NMDA-channel activity appeared to track [Na⁺]_i over the range 0–40 mM.

Intracellular [Na⁺]_i is increased markedly by activating ionotropic glutamate receptors⁴. To determine whether Na⁺ influx during activation of these receptors modulates NMDA channels, we probed channel activity by enclosing single channels within a patch pipette attached to the cell and activating surrounding NMDA receptors or non-NMDA receptors by applying selective agonists to the bath solution. The bath-applied agonists produced a stable level of cell depolarization but did not affect the conductance of the NMDA channels in the patches (Fig. 2e). Thus, we avoided voltage-dependent changes in NMDA-channel activity⁹ by adjusting the potential of the patch to maintain a transmembrane potential that was 70 mV negative to the NMDA-channel reversal potential.

To activate NMDA receptors outside the patch, we bath-applied the selective agonist L-aspartate. This significantly increased *P*_o, *t*_o and the lengths of bursts and clusters of NMDA channels within the patches (Fig. 2a–e). The increase in NMDA single-channel activity was prevented by applying the NMDA-receptor antagonist D(-)-amino-7-phosphonovaleric acid (AP5) in the bath together with L-aspartate (Fig. 2d); bath application of AP5 alone had no effect (*P*_o was 99 ± 6% control; *n* = 6 patches). Moreover, bath-applying K⁺ (30 mM), which produced depolarization comparable with that caused by L-aspartate (20 ± 9 mV, *n* = 8 cells, compared with 20 ± 4 mV, *n* = 6 cells, respectively), did not alter NMDA single-channel activity (Fig. 2d). Together these results indicate that there is a positive crosstalk between NMDA receptors through a diffusible intracellular factor.

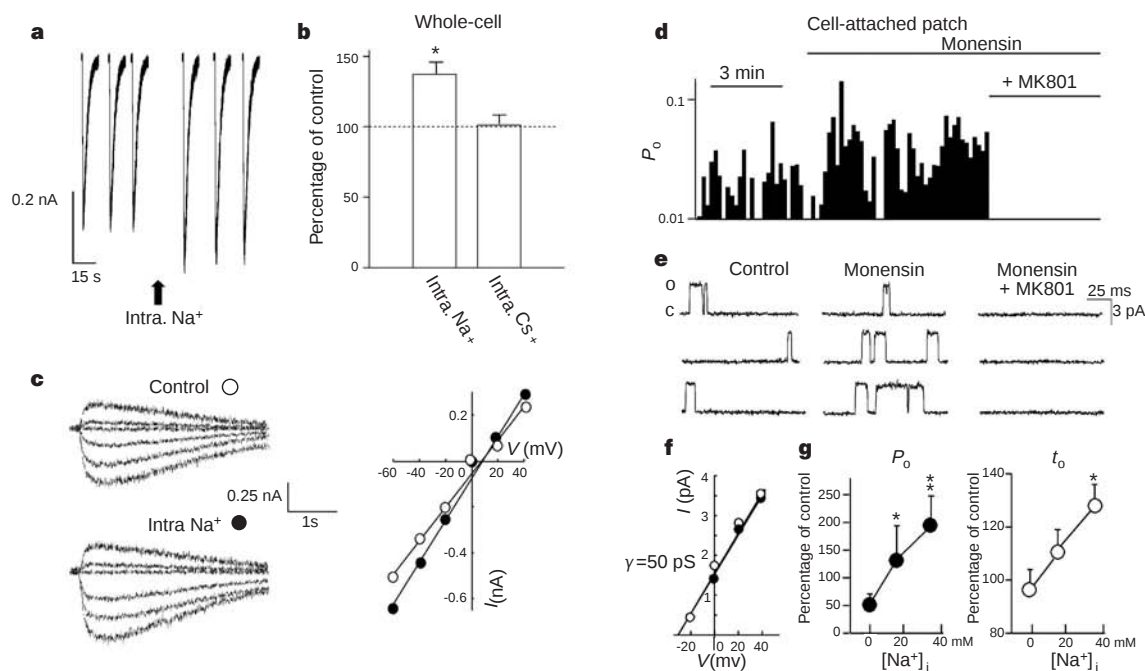


Figure 1 NMDA-channel activity is regulated by intracellular Na^+ . **a**, Current traces from a representative recording show individual whole-cell responses evoked by applying NMDA ($100 \mu\text{M}$) before and during intracellular perfusion with 50 mM Na^+ solution (arrow). **b**, Effects of intracellular application of solutions containing 50 mM Na^+ ($n = 4$ neurons) or 50 mM Cs^+ ($n = 4$ neurons). Values are mean $\pm$ s.e.m. The dotted line indicates the control level before intracellular perfusion. **c**, Graph showing the current-voltage (I - V) relationship of the peak of the NMDA current before (open circles) and during (filled circles) the intracellular application of Na^+ . The lines through the data points are least-squares regression lines. Individual NMDA currents at the different membrane potentials are shown to the left of the graph. **d-g**, Effects of monensin on NMDA single-

channel activity recorded in cell-attached patches. **d**, A continuous record of NMDA-channel P_o from a neuron bathed with extracellular solution containing 50 mM Na^+ . Monensin ($10 \mu\text{M}$) and monensin + MK801 ($2 \mu\text{M}$) were applied in the bath during the periods indicated. P_o was calculated in 10-s bins. **e**, Single-channel currents before (control; left traces) and during application of monensin (middle traces) or monensin + MK801 ($2 \mu\text{M}$; right traces); c, closed level; o, open level. Each set of three traces is a continuous recording. **f**, I - V relationship of the channel before (open circles) and during (filled circles) monensin application. **g**, P_o and t_o plotted against $[\text{Na}^+]_i$; $n = 6$ patches for each point. (Single asterisks, $P < 0.05$; double asterisks, $P < 0.01$; Mann-Whitney U-test.)

NMDA receptors are permeable to Ca^{2+} , as well as to Na^+ , and therefore the activation of the remote NMDA receptors might have resulted from an influx of Ca^{2+} . However, bath application of L-aspartate increased NMDA single-channel activity in the patches even when the bath solution had no added Ca^{2+} and contained the Ca^{2+} chelator BAPTA (1 mM) (Fig. 2f, g). In contrast, removal of Na^+ from the bathing solution prevented the potentiation of single-channel activity (Fig. 2g). When Na^+ -free bathing solution was used, L-aspartate significantly reduced the P_o of the NMDA channels, which may be attributable to a reduction in $[\text{Na}^+]_i$ induced by Na^+ efflux through NMDA receptors. Similarly, the activation of non-NMDA receptors, by bath application of kainate, increased NMDA-channel activity recorded in the cell-attached patches in bath solution containing Na^+ (Fig. 3a, b) and reduced NMDA-channel activity in the presence of Na^+ -free bath solution (Fig. 3b). Thus, Na^+ influx through either subtype of glutamate receptor potentiated the activity of NMDA channels.

A major route for Na^+ entry into neurons is through voltage-gated channels that are permeable to Na^+ , such as those that underlie the generation of action potentials¹. To determine whether influx of Na^+ through these voltage-gated Na^+ channels is sufficient to affect NMDA-channel function, we applied veratridine; this removes the inactivation of these channels¹⁰, thereby allowing them to remain open and so producing a sustained level of cell depolarization. Bath application of veratridine significantly increased NMDA-channel activity in cell-attached patches (Fig. 3c, d), but had no effect on single-channel conductance. The effects of veratridine were prevented when it was applied with tetrodotoxin (TTX; Fig. 3d). Thus, influx of Na^+ through TTX-sensitive voltage-

gated Na^+ channels is sufficient to produce potentiation of NMDA-channel activity.

The observations from the steady-state single-channel recordings indicate that individual NMDA-receptor activations are regulated by intracellular Na^+ . Because synaptic responses to NMDA are produced by single activations of the receptors¹¹, we determined whether synaptically activated NMDA currents are potentiated by raising $[\text{Na}^+]_i$. We examined miniature excitatory postsynaptic currents (mEPSCs) and found that applying Na^+ into the neurons through the patch pipette significantly increased the NMDA-receptor-mediated component of the mEPSCs (Fig. 4a, b). In contrast, raising $[\text{Na}^+]_i$ had no effect on the non-NMDA-receptor component. Applying Cs^+ intracellularly did not affect either NMDA-receptor-mediated or non-NMDA-receptor-mediated components of the mEPSCs (Fig. 4b). Thus, increasing $[\text{Na}^+]_i$ selectively enhanced synaptic transmission mediated by NMDA receptors but not non-NMDA receptors.

To study the regulation of NMDA channels by Na^+ in isolation from the cells, we used excised patches in the inside-out configuration. In contrast to effects in cell-attached recordings, increasing $[\text{Na}^+]$ to 50 mM on the cytoplasmic face of inside-out patches had no effect on NMDA-channel activity: P_o was $93 \pm 12\%$ of control and t_o was $92 \pm 7\%$ of control ($n = 6$ patches). Thus, there might be biochemical contingencies regulating the actions of Na^+ on NMDA channels, one such contingency being protein phosphorylation. When recording NMDA-channel activity in the cell-attached patch configuration, we found that administering the broad-spectrum protein-kinase inhibitor staurosporine prevented the increase in channel activity when monensin was used to raise $[\text{Na}^+]_i$ in the

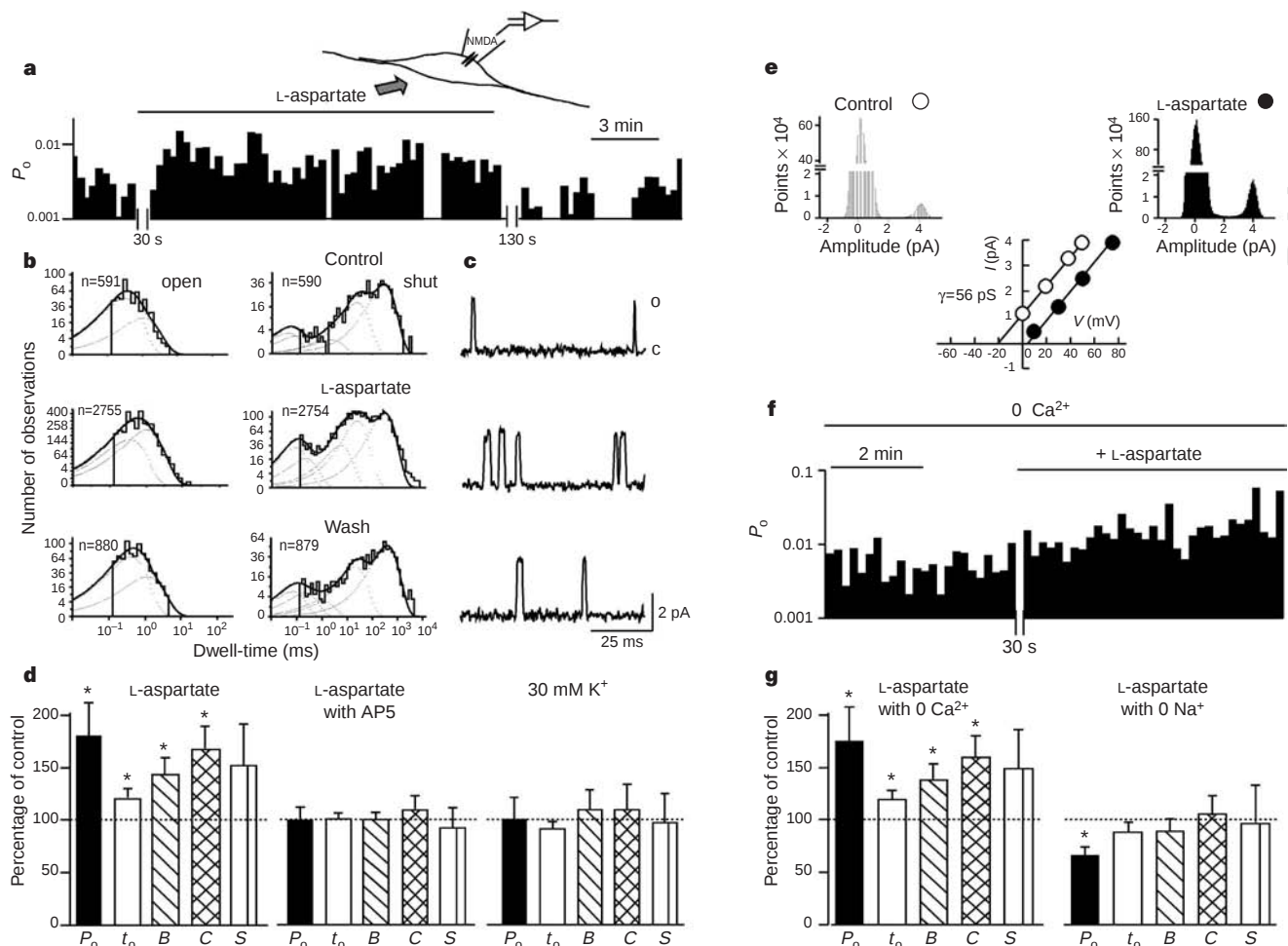


Figure 2 NMDA single-channel activity is potentiated by Na^+ influx through remote NMDA receptors. **a**, A record of NMDA-channel P_o before, during and after bath application of L-aspartate ($500 \mu\text{M}$). The recording is discontinuous, and includes breaks needed to adjust the patch potential to 70 mV from the NMDA-channel reversal potential. **b**, Dwell-time histograms of open and shut times compiled before (upper panel), during (middle panel) and after (lower panel) bath application of L-aspartate. The solid curve shows the Marquardt least-squares fit for the histogram; dashed curves indicate individual exponential components. **c**, Representative single-channel current traces before, during and after application of L-aspartate. **d**, Effects of bath-applied L-aspartate (8 patches), L-aspartate in the presence of AP5 ($20 \mu\text{M}$; 4 patches), or K^+ (30 mM ; 6 patches). B, C and S are

lengths of bursts, clusters and superclusters, respectively. **e**, All-point histograms of a recording before and during bath application of L-aspartate with the patch potential 70 mV from the reversal potential. The graph shows the I - V relationship for single-channel currents before (open circles) or during (filled circles) L-aspartate application. The lines are least-squares regression lines. **f**, A record of NMDA-channel P_o during cell-attached recording when the bath solution contained BAPTA and no added Ca^{2+} . L-aspartate ($500 \mu\text{M}$) was applied to the bath solution during the period indicated. The break during the recording was for adjusting patch potential. **g**, Effects on NMDA single-channel activity of L-aspartate applied in Ca^{2+} -free ($n = 6$ patches) and Na^+ -free ($n = 6$ patches) bathing solution. (Single asterisks, $P < 0.05$; Wilcoxon test.)

presence of 50 mM extracellular Na^+ (P_o $85 \pm 20\%$ of control, t_o $98 \pm 4\%$ of control; $n = 6$ patches) but it did not affect the monensin-induced rise in $[\text{Na}^+]_i$ ($[\text{Na}^+]_i$ increased by $30 \pm 5 \text{ mM}$; $n = 7$ cells).

Thus, we questioned whether enhancing phosphorylation may allow Na^+ to affect NMDA channels in inside-out patches. Because tyrosine phosphorylation by Src upregulates NMDA-channel function in inside-out patches¹², we examined the effects of activating Src kinases with the phosphopeptide EPQ(pY)EEIPIA¹³ (Fig. 5). Application of Na^+ (30 or 50 mM) to the cytoplasmic face of the patches pretreated with EPQ(pY)EEIPIA caused a reversible increase in NMDA-channel activity (Fig. 5b, c; Supplementary Information). In contrast, in patches treated with the dephosphorylated peptide, EPQYEEIPIA, which does not activate Src kinases¹³, 50 mM Na^+ had no effect on NMDA-channel activity. To determine whether NMDA channels in inside-out patches could be affected without enhancing tyrosine kinase activity, we applied 80 mM Na^+ to the cytoplasmic face of patches not treated with EPQ(pY)EEIPIA;

this produced an increase in channel activity (Fig. 5c). Together, these results indicate that Na^+ may be acting directly on the channel or on a closely associated regulatory site and, moreover, that the sensitivity of NMDA channels to Na^+ is set by a Src kinase associated with the channel

Our results show that the activity of NMDA receptors is modulated by intracellular $[\text{Na}^+]_i$. Na^+ is the principal charge-carrying ion responsible for producing membrane depolarization during action potentials and excitatory synaptic transmission. Our results indicate that the functional role of Na^+ in neurons is, however, not limited to that of a charge carrier. Instead, our findings show that once Na^+ enters postsynaptic neurons it may function as a signalling ion. The effect of raising $[\text{Na}^+]_i$ on NMDA receptors is opposite to that of raising intracellular $[\text{Ca}^{2+}]_i$, which depresses channel function^{3,14-16}. During high levels of neuronal firing, $[\text{Na}^+]_i$ increases to 15–20 mM in the neuronal soma³, concentrations we found to increase NMDA-channel activity, and to higher levels in the dendrites². Thus, the Na^+ -dependent upregulation of NMDA receptors seems to be an

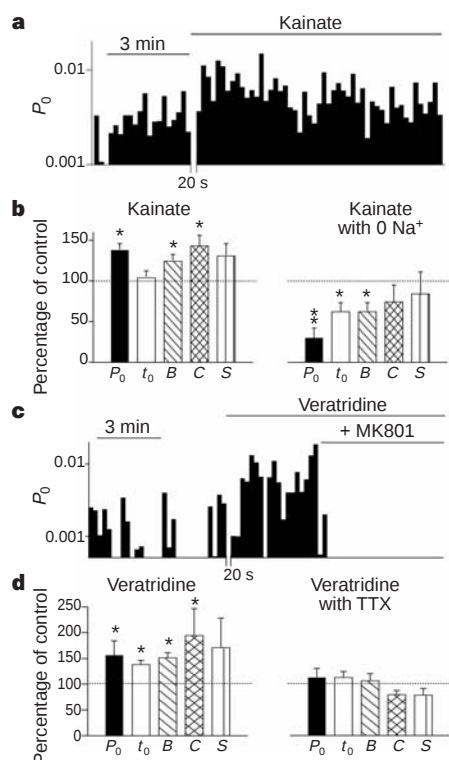


Figure 3 NMDA single-channel activity is upregulated by activating non-NMDA receptors and voltage-gated Na^+ channels. **a**, A record of NMDA-channel P_o before and during bath application of kainate (1 mM). The break was for adjustment of patch potential. **b**, Effects of kainate on NMDA single-channel activity in cell-attached patches on neurons bathed with standard ($n = 6$ patches) and Na^+ -free ($n = 4$ patches) extracellular solution. **c**, Recording of NMDA-channel P_o before and during bath application of veratridine (10 μM) or veratridine + MK801 (2 μM). **d**, Effects of bath application of veratridine alone ($n = 6$ patches) or veratridine plus TTX (1 μM ; $n = 6$ patches). (Single asterisks, $P < 0.05$; double asterisks, $P < 0.01$; Wilcoxon test or paired t -test.)

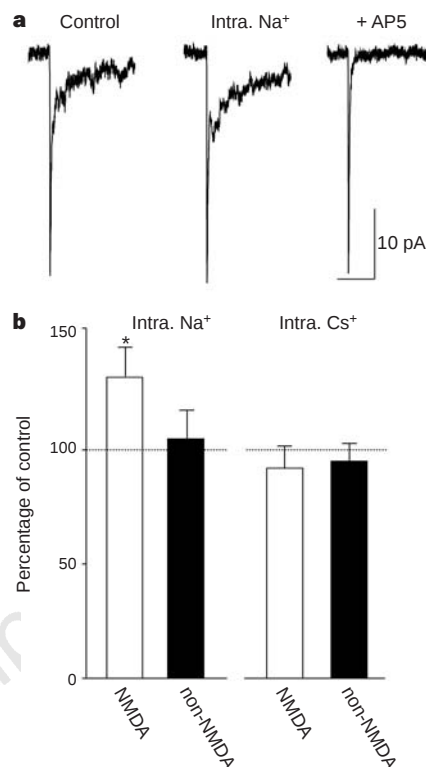


Figure 4 Synaptic NMDA receptors but not non-NMDA receptors are regulated by intracellularly applied Na^+ . **a**, Representative recordings of averaged mEPSCs before (left) and during the intracellular application of Na^+ by perfusing the pipette with solution containing 50 mM Na^+ before (middle) and during (right) bath application of AP5. **b**, Effects of the intracellular application of Na^+ ($n = 6$ neurons) or Cs^+ (50 mM; $n = 6$ neurons) on the charge of the NMDA and non-NMDA components. (Single asterisk, $P < 0.05$; Mann-Whitney U-test.)

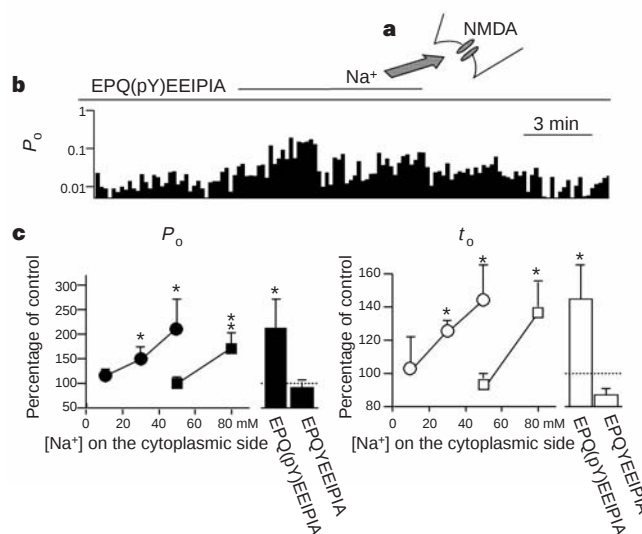


Figure 5 Raising the Na^+ concentration on the cytoplasmic face of excised patches increases NMDA-channel activity during activation of Src-family kinases. **a**, Illustration of the use of inside-out patches. **b**, A continuous record of NMDA-channel P_o . EPQ(pY)EEIPIA was applied throughout the recording period, starting 3 min before the beginning of the record. Na^+ (50 mM) was applied to the cytoplasmic face during the period indicated. In **c**, the line graphs show mean changes in P_o or t_o at varying $[\text{Na}^+]$ on the cytoplasmic side of the membrane in the presence (circles) or

the absence (squares) of EPQ(pY)EEIPIA. In the presence of the activator peptide, Na^+ was tested at 10 mM ($n = 3$ patches), 30 mM ($n = 6$ patches) or 50 mM ($n = 4$ patches). Without activator peptide, Na^+ concentrations were 50 mM ($n = 6$ patches) or 80 mM ($n = 4$ patches). (Single asterisks, $P < 0.05$; double asterisks, $P < 0.01$; Wilcoxon test or paired t -test.) The bar graphs show mean changes in P_o or t_o upon application of 50 mM Na^+ in the presence of EPQ(pY)EEIPIA or of EPQYEEIPIA ($n = 4$ patches). (Single asterisks, $P < 0.05$; Mann-Whitney U-test.)

unusual mechanism for amplifying NMDA-receptor-mediated synaptic responses at the same time as high levels of firing occur. High levels of neuronal firing are associated with increased activity of Src kinase¹⁷, and thus NMDA channels are sensitized to Na⁺ under such circumstances. Single action potentials or short trains produce smaller changes² in [Na⁺]_i, which indicates that NMDA-channel function may be tuned to the level of neuronal activity. Moreover, neurons express a diversity of Na⁺-permeable channels which are activated by several stimuli^{18–22}. Maintaining intracellular [Na⁺] at the resting level requires considerable expenditure of energy and, consequently, [Na⁺]_i rises when neurons are metabolically compromised²³. Thus, [Na⁺]_i integrates a wide range of physiological and pathophysiological stimuli and provides a mechanism for modulating NMDA-channel activity during various physiological and pathological events in the central nervous system. □

Methods

Tissue culture and whole-cell recording. Primary cultures from spinal dorsal horn or hippocampus were prepared from fetal Wistar rats (embryonic day 17–19) as described²⁴. For whole-cell recordings, the cultures were bathed in extracellular solution²⁵ supplemented with TTX (1 μM). For recording mEPSCs, the extracellular solution was further supplemented with bicuculline (10 μM) and strychnine (10 μM). Recording pipettes²⁵ contained intracellular solution composed of (in mM): KCl, 140; CaCl₂, 1; BAPTA, 10; HEPES, 10; MgCl₂, 2; K-ATP, 4; pH 7.25, osmolarity 300–310 mosM. For the high [Na⁺] or [Cs⁺] intracellular solutions, KCl was replaced equimolarly by NaCl or CsCl, respectively. Cs⁺ was used because the Na⁺:Cs⁺ permeability ratio of NMDA channels is ~1 (ref. 26). The solutions were perfused intracellularly by means of a pipette inserted into the recording electrode²⁷. Recordings were made under voltage-clamp conditions and the holding potential was –60 mV, except where indicated. Whole-cell NMDA currents were evoked by sodium NMDA (100 μM) or sodium L-aspartate (500 μM); these compounds were dissolved in extracellular solution and applied by pressure²⁵.

Visualization of neurons. For recordings, the cultures were placed in a chamber on an inverted microscope (Nikon) and a magnified image of the field was displayed on a television monitor so that cell shape could be monitored, and the diameter of the soma and major processes could be measured at regular intervals during the recording period. With the bath solutions used, none of the experimental manipulations produced changes in size or shape of the cell soma or main processes. Changes in cell morphology were, however, easily detected by changing the osmolarity of the extracellular solution. Thus, the experimental manoeuvres used did not seem to cause cell stretching, which might have produced changes in NMDA-channel activity²⁸.

Single-channel recording. Single-channel recording and methods for analysing single openings and groupings of openings, as well as criteria used to ensure that recordings were from NMDA channels, have been described^{12,29}. NMDA-receptor-mediated single-channel currents were evoked by including in the recording pipette a standard extracellular solution¹² supplemented with 10 μM NMDA and 3 μM glycine. With recordings in cell-attached mode, Na⁺ in the bathing solution was replaced wholly or in part by N-methyl-D-glucamine as necessary. TTX (1 μM) was included in the bath solution in experiments in which glutamate-receptor agonists or high [K⁺] were applied. To study the effects of veratridine (10 μM), synaptic transmission was blocked by using extracellular bath solution with no added Ca²⁺ plus BAPTA (1 mM) and Mg²⁺ (5 mM), and the main ionotropic neurotransmitter receptors were blocked by including AP5 (100 μM), 6,7-dinitroquinoxaline-2,3-dione (DNQX) (10 μM), bicuculline (10 μM) and strychnine (10 μM). During recordings with the inside-out patch configuration, an intracellular solution¹² was applied to the cytoplasmic face of the membrane. The Cs⁺ and Na⁺ concentrations in the intracellular solution were adjusted as required and the peptides EPQ(pY)EEIPIA or EPQYEEIPIA were added as necessary. Patches were treated with the peptides for 5–8 min to allow channel activity to reach a stable level.

NMDA single-channel activity was studied with a patch potential that was 70 mV from the reversal potential. Channel activity during the entire period of applying test substances (typically 4–10 min), beginning at the start of the

application or, in cases in which membrane potential changed, beginning when membrane potential had stabilized, was compared with that for a similar period immediately preceding the application.

mEPSC recording and analysis. Miniature EPSCs were recorded in whole-cell mode and were detected semi-automatically off-line and analysed after alignment of the rising edge, as described¹². mEPSCs were averaged for 3–10-min periods before and during intracellular perfusion. Averages had to include at least 30 traces. The mEPSCs studied are glutamatergic, consisting of a rapid, transient component mediated by non-NMDA receptors and a slower component mediated by NMDA receptors^{12,30}. The NMDA-receptor-mediated component was calculated by digitally subtracting, from the averaged mEPSC, a current decaying at a single exponential rate that was equal to that of the non-NMDA component¹². We calculated charge by integrating currents during NMDA or non-NMDA components of mEPSCs.

Intracellular [Na⁺] measurement. [Na⁺]_i was measured by means of the Na⁺-sensitive fluorescent dye SBFI⁷, using a 340/380-nm filter set and single-photon-counting methods used^{24,25} for recording [Ca²⁺]_i with fura-2. Cells were loaded either with SBFI-AM (20 μM) applied by bath for 90–120 min at 22 °C or with SBFI-free acid (100 μM) in intracellular solution applied through the recording pipette. For each cell, 340/380-nm fluorescence ratios were calibrated *in situ* with bath application of monensin (10 μM) and ouabain (1.5 mM) to equilibrate intracellular with extracellular [Na⁺]. Calibration was done using two (0, 50 mM) or three (0, 20, 50 mM) Na_o⁺ concentrations.

All chemicals used were from Sigma except for BAPTA, SBFI and monensin, which were from Molecular Probes, and MK801, which was from MSD (Harlow). For further information on Methods and Figs 1, 2 and 5 see Supplementary Information.

Received 4 August; accepted 14 October 1998.

- Hille, B. *Ionic Channels of Excitable Membranes* (Sinauer, Sunderland, 1992).
- Jaffe, D. B. et al. The spread of Na⁺ spikes determines the pattern of dendritic Ca²⁺ entry into hippocampal neurons. *Nature* **357**, 244–246 (1992).
- Rose, C. R. & Ransom, B. R. Regulation of intracellular sodium in cultured rat hippocampal neurons. *J. Physiol. (Lond.)* **499**, 573–587 (1997).
- Callaway, J. C. & Ross, W. N. Spatial distribution of synaptically activated sodium concentration changes in cerebellar Purkinje neurons. *J. Neurophysiol.* **77**, 145–152 (1997).
- McBain, C. J. & Mayer, M. L. N-methyl-D-aspartic acid receptor structure and function. *Physiol. Rev.* **74**, 723–760 (1994).
- Sandeaux, R., Sandeaux, J., Gavach, C. & Brun, B. Transport of Na⁺ by monensin across bimolecular lipid membranes. *Biochim. Biophys. Acta* **684**, 127–132 (1982).
- Harootunian, A. T., Kao, J. P., Eckert, B. K. & Tsien, R. Y. Fluorescence ratio imaging of cytosolic free Na⁺ in individual fibroblasts and lymphocytes. *J. Biol. Chem.* **264**, 19458–19467 (1989).
- Wong, E. H. & Kemp, J. A. Sites for antagonism on the N-methyl-D-aspartate receptor channel complex. *Annu. Rev. Pharmacol. Toxicol.* **31**, 401–425 (1991).
- Nowak, L. M. & Wright, J. M. Slow voltage-dependent changes in channel open-state probability underlie hysteresis of NMDA responses in Mg(2+)-free solutions. *Neuron* **8**, 181–187 (1992).
- Strichartz, G., Rando, T. & Wang, G. K. An integrated view of the molecular toxicology of sodium channel gating in excitable cells. *Annu. Rev. Neurosci.* **10**, 237–267 (1987).
- Edmonds, B. & Colquhoun, D. Mechanisms of activation of glutamate receptors and the time course of excitatory synaptic currents. *Annu. Rev. Physiol.* **57**, 495–519 (1995).
- Yu, X. M., Askalan, R., Keil, G. J. I. & Salter, M. W. NMDA channel regulation by channel-associated protein tyrosine kinase Src. *Science* **275**, 674–678 (1997).
- Liu, X. et al. Regulation of c-Src tyrosine kinase activity by the Src SH2 domain. *Oncogene* **8**, 1119–1126 (1993).
- Legendre, P., Rosenmund, C. & Westbrook, G. L. Inactivation of NMDA channels in cultured hippocampal neurons by intracellular calcium. *J. Neurosci.* **13**, 674–684 (1993).
- Rosenmund, C. & Westbrook, G. L. Calcium-induced actin depolymerization reduces NMDA channel activity. *Neuron* **10**, 805–814 (1993).
- Medina, I., Filippova, N., Bakhramov, A. & Bregestovski, P. Calcium-induced inactivation of NMDA receptor-channels evolves independently of run-down in cultured rat brain neurons. *J. Physiol. (Lond.)* **495**, 411–427 (1996).
- Lu, Y. M., Roder, J. C., Davidow, J. & Salter, M. W. Src activation in the induction of long-term potentiation in CA1 hippocampal neurons. *Science* **279**, 1363–1368 (1998).
- Crill, W. E. Persistent sodium current in mammalian central neurons. *Annu. Rev. Physiol.* **58**, 349–362 (1996).
- Parri, H. R. & Crunelli, V. Sodium current in rat and cat thalamocortical neurons: role of a non-inactivating component in tonic and burst firing. *J. Neurosci.* **18**, 854–867 (1998).
- Pape, H. C. Queer current and pacemaker: the hyperpolarization-activated cation current in neurons. *Annu. Rev. Physiol.* **58**, 299–327 (1996).
- Congar, P., Leinekugel, X., Ben-Ari, Y. & Crepel, V. A long-lasting calcium-activated nonselective cationic current is generated by synaptic stimulation or exogenous activation of group I metabotropic glutamate receptors in CA1 pyramidal neurons. *J. Neurosci.* **17**, 5366–5379 (1997).
- Xiong, Z., Lu, W. & MacDonald, J. F. Extracellular calcium sensed by a novel cation channel in hippocampal neurons. *Proc. Natl Acad. Sci. USA* **94**, 7012–7017 (1997).
- Chidekel, A. S., Friedman, J. E. & Haddad, G. G. Anoxia-induced neuronal injury: role of Na⁺ entry and Na⁺-dependent transport. *Exp. Neurol.* **146**, 403–413 (1997).
- Salter, M. W. & Hicks, J. L. ATP-evoked increases in intracellular calcium in cultured neurones and glia from the dorsal spinal cord. *J. Neurosci.* **14**, 1563–1575 (1994).
- Wang, Y. T. & Salter, M. W. Regulation of NMDA receptors by tyrosine kinases and phosphatases. *Nature* **369**, 233–235 (1994).
- Mayer, M. L. & Westbrook, G. L. The physiology of excitatory amino acids in the vertebrate central nervous system. *Prog. Neurobiol.* **28**, 197–276 (1987).

27. Wang, L.-Y., Salter, M. W. & MacDonald, J. F. Regulation of kainate receptors by cAMP-dependent protein kinase and phosphates. *Science* **253**, 1132–1135 (1991).
28. Paoletti, P. & Ascher, P. Mechanosensitivity of NMDA receptors in cultured mouse central neurons. *Neuron* **13**, 645–655 (1994).
29. Wang, Y. T., Yu, X. M. & Salter, M. W. Ca²⁺-independent reduction of N-methyl-D-aspartate channel activity by protein tyrosine phosphatase. *Proc. Natl. Acad. Sci. USA* **93**, 1721–1725 (1996).
30. Bekkers, J. M. & Stevens, C. F. NMDA and non-NMDA receptors are co-localized at individual excitatory synapses in cultured rat hippocampus. *Nature* **341**, 230–233 (1989).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. We thank J. F. MacDonald, Y. T. Wang, L. Y. Wang and Y. De Koninck for comments on the manuscript; Y. De Koninck for software used in analysis; and J. L. Hicks for preparing and maintaining cell cultures and for technical assistance. This work was supported by a grant from the Medical Research Council of Canada to M.W.S. and by the Nicole Fealdman Memorial Fund. M.W.S. is an MRC Scientist and X.-M.Y. is an MRC Scholar.

Correspondence and requests for materials should be addressed to M.W.S. (e-mail: mike.salter@utoronto.ca).

Epac is a Rap1 guanine-nucleotide-exchange factor directly activated by cyclic AMP

Johan de Rooij^{*,†}, Fried J. T. Zwartkruis^{*,†}, Mark H. G. Verheijen^{*,‡}, Robbert H. Cool[§], Sebastian M. B. Nijman^{*}, Alfred Wittinghofer[§] & Johannes L. Bos^{*}

^{*} Laboratory for Physiological Chemistry and Centre for Biomedical Genetics, Utrecht University, Universiteitsweg 100, 3584 CG Utrecht, The Netherlands
[†] Hubrecht Laboratory, Netherlands Institute for Developmental Biology, 3584 CT Utrecht, The Netherlands
[§] Max Planck Institute für Molekulare Physiologie, 44139 Dortmund, Germany
[‡] These authors contributed equally to this work.

Rap1 is a small, Ras-like GTPase that was first identified as a protein that could suppress the oncogenic transformation of cells by Ras¹. Rap1 is activated by several extracellular stimuli^{2–7} and may be involved in cellular processes such as cell proliferation⁸, cell differentiation⁴, T-cell anergy² and platelet activation⁷. At least three different second messengers, namely diacylglycerol, calcium and cyclic AMP^{5–7,9}, are able to activate Rap1 by promoting its release of the guanine nucleotide GDP and its binding to GTP. Here we report that activation of Rap1 by forskolin and cAMP occurs independently of protein kinase A (also known as cAMP-activated protein kinase). We have cloned the gene encoding a guanine-nucleotide-exchange factor (GEF) which we have named Epac (exchange protein directly activated by cAMP). This protein contains a cAMP-binding site and a domain that is homologous to domains of known GEFs for Ras and Rap1. Epac binds cAMP *in vitro* and exhibits *in vivo* and *in vitro* GEF activity towards Rap1. cAMP strongly induces the GEF activity of Epac towards Rap1 both *in vivo* and *in vitro*. We conclude that Epac is a GEF for Rap1 that is regulated directly by cAMP and that Epac is a new target protein for cAMP.

Treatment of CHO cells with forskolin, a drug that induces cAMP formation by activating adenylate cyclase, results in an increase in the amount of the active, GTP-bound form of Rap1 in a concentration-dependent manner (Fig. 1a). This result confirms previous observations that cAMP is able to activate Rap1 in various cell types^{6,9,10}. To determine whether forskolin-induced Rap1 activation is mediated by protein kinase A (PKA), we compared forskolin-induced Rap1 activation in parental CHO10001 and PKA-mutant CHO10248 cell lines. CHO10248 cells express a mutant regulatory subunit of PKA which binds cAMP with a much lower affinity¹¹.

Rap1 was activated equally well in both cell lines. A comparison of the levels of PKA activity induced by forskolin in both cell lines showed that PKA activity was induced in the CHO10001 cells. In contrast, in CHO10248 cells, the basal level of PKA activity was reduced and after forskolin stimulation PKA activity was induced to a level that is similar to the basal activity of the parental cell line. To exclude the possibility that this increase in PKA activity is sufficient to activate Rap1, we treated CHO10248 cells with the PKA inhibitor H89. Although this treatment completely abolished PKA activity, it did not affect forskolin-induced Rap1 activity (Fig. 1b). Forskolin, 8-bromoadenosine 3',5'-cyclic monophosphate (8-Br-cAMP) and 8-chlorophenylthio-cAMP (CPT-cAMP; two membrane-permeable cAMP analogues), but not inactive dideoxy-forskolin, activated Rap1 in CHO10248 cells, showing that cAMP is responsible for the observed activation of Rap1 (Fig. 1c). In Rat1 cells, forskolin-induced GTP-bound Rap1 is also insensitive to H89 (Fig. 1d). We conclude that Rap1 is activated by cAMP independently of PKA.

As there is a rasGEF that is regulated by direct binding of second messengers, namely Ca²⁺ and diacylglycerol¹², we searched databases for proteins with sequence homology both to GEFs for Ras and Rap1 and to cAMP-binding sites. We identified a genomic sequence with the characteristics of a gene encoding such a protein. We used oligonucleotides based on the presumed exons in this sequence to isolate, by reverse transcription with polymerase chain reaction (RT-PCR) of messenger RNA from human sources, a complementary DNA encoding a 881-amino-acid protein which we named Epac (Fig. 2a).

The putative cAMP-binding site of Epac shares significant sequence identity with the slow and fast cAMP-binding sites of the two regulatory subunits of PKA (Fig. 2b). Epac also shares similarity with the cyclic-nucleotide-gated channels, such as those present in olfactory cells¹³. Epac has a characteristic alanine in the

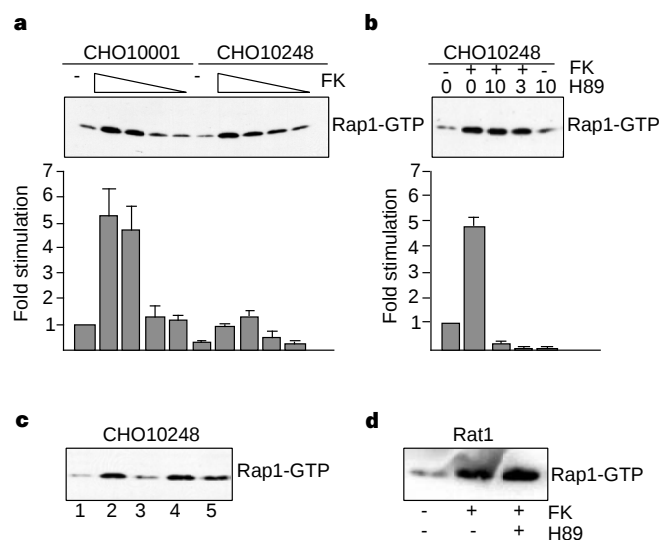


Figure 1 cAMP-induced Rap1 activation is not mediated by PKA. **a**, CHO10001 and CHO10248 cells were incubated with forskolin (FK) (10, 2, 0.2 or 0.05 μ M, from left to right) for 10 min and the amount of GTP-bound Rap1 (Rap1-GTP) was analysed^{6,7}. PKA activity was also determined (means $\pm$ s.e.m., $n = 3$)²⁴. **b**, CHO10248 cells were incubated with the PKA inhibitor H89 (10 or 3 μ M) for 30 min and stimulated with forskolin (10 μ M) for 10 min. The amounts of Rap1-GTP and PKA activity were determined. **c**, CHO10248 cells were stimulated with forskolin (10 μ M, lane 2), dideoxy-forskolin (10 μ M, lane 3), 8-Br-cAMP (500 μ M, lane 4) or 8-CTP-cAMP (150 μ M, lane 5), and the amount of Rap1-GTP was determined. Lane 1, control. **d**, Rat1 cells were incubated with forskolin (10 μ M, 10 min) or with H89 (10 μ M, 30 min) and forskolin (10 μ M, 10 min), and the amount of Rap1-GTP was determined.

cAMP-binding pocket; this alanine is also present in the cAMP-binding sites of PKA, whereas the cyclic-nucleotide-gated channels have a threonine at this position. This alanine is considered to be responsible for the specificity of binding of cAMP rather than cGMP¹⁴. The GEF-homology domain of Epac shares significant homology with C3G, the only Rap1-specific GEF described so far¹⁵ (Fig. 2c). Epac also contains a Ras exchange motif (REM) and a Dishevelled, Egl-10, Pleckstrin (DEP) domain. The REM region has been found in all known Ras- and Rap1-specific GEFs and may be important in the stabilization of the GEF structure¹⁶. The DEP domain¹⁷ may be involved in membrane attachment. For example, the DEP domain in *Drosophila* Dishevelled is necessary and sufficient for Frizzled-mediated relocation of Dishevelled to the membrane¹⁸. Northern blot analysis showed that Epac is expressed in all human tissues tested, but is particularly abundant in kidney and heart (Fig. 2d). Epac mRNA could also be detected in fibroblast cell lines such as Rat1 cells (Fig. 2e), which are responsive to cAMP-induced Rap1 activation (Fig. 1d).

To investigate whether the cAMP-binding domain of Epac can bind cAMP, we incubated fusion proteins (Fig. 3a), containing glutathione-S-transferase (GST) and Epac, the cAMP-binding domain of Epac (Epac-RD) or the catalytic domain of Epac

(Epac-ΔcAMP) bound to glutathione beads, with radiolabelled cAMP. Both Epac and Epac-RD, but not Epac-ΔcAMP, GST alone or an unrelated protein (RalGDS-RBD), retained radiolabelled cAMP (Fig. 3b). This binding could be eliminated by competition with an excess of unlabelled cAMP. As expected, the regulatory subunit Iα of PKA (RIα) also binds cAMP under these conditions. A comparison of cAMP binding to Epac and RIα showed that the concentrations of cAMP necessary for half-maximal binding to these proteins are in the same order of magnitude, indicating similar affinity (Fig. 3c). These results indicate that cAMP interacts directly with Epac.

To determine whether Epac is a Rap1-specific GEF and whether it responds to cAMP, we used NIH3T3-A14 cells. This cell line contains Epac mRNA (Fig. 2e), but stimulation of these cells with forskolin or 8-Br-cAMP resulted in no, or only a small, increase in the amount of endogenous or ectopically expressed Rap1 (Fig. 4a, c). However, co-transfection of Rap1 with Epac (Fig. 4b) led to a rise in the basal level of GTP-bound epitope-tagged Rap1 and, more important, Rap1 activity increased markedly after stimulation with forskolin (Fig. 4c, left). Ectopic expression of Epac also rescued 8-Br-cAMP-induced Rap1 activation (Fig. 4c, right). Epac-ΔcAMP also induced Rap1 activation but this activation was

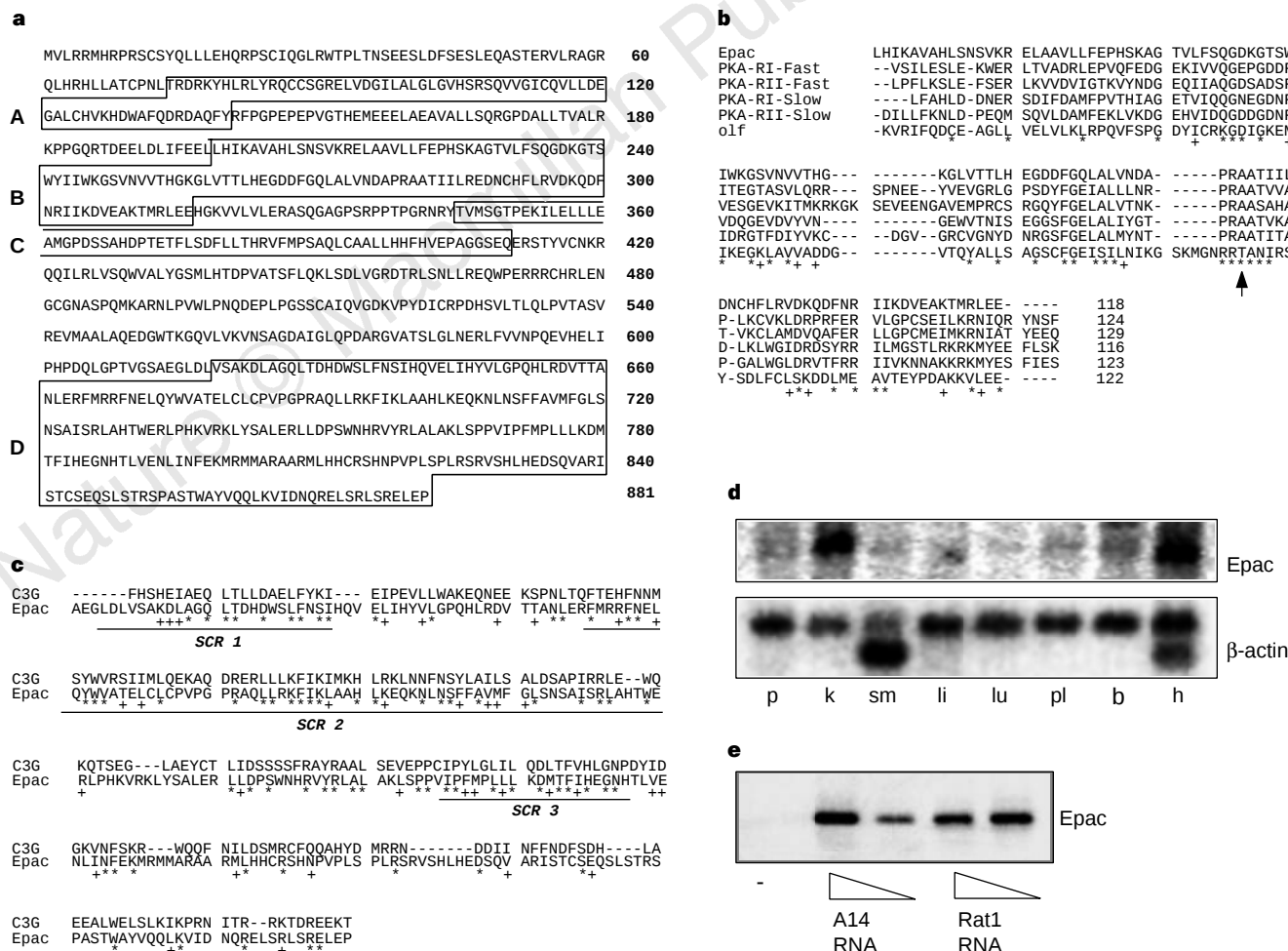


Figure 2 Sequence of Epac. **a**, Protein sequence of Epac with domains homologous to other protein domains being boxed. A, DEP domain¹⁷; B, cAMP-binding domain; C, REM domain; D, catalytic domain. **b**, Comparison of the cAMP-binding domains of Epac, PKA-RI and -RII and a cyclic-nucleotide-gated olfactory channel. Amino acids of Epac that are identical to residues of at least two other proteins are marked with an asterisk and conserved changes in Epac are marked with a plus symbol. The arrowhead indicates the critical alanine

residue. **c**, Comparison of the catalytic domain of Epac with that of C3G, the only previously known Rap1-specific GEF. Strongly conserved regions of GEFs for Ras and Rap1 are indicated (SCR 1-3). **d**, Expression of Epac mRNA in human pancreas (p), kidney (k), skeletal muscle (sm), liver (li), lung (lu), placenta (pl), brain (b) and heart (h). **e**, Expression of Epac mRNA in NIH3T3-A14 and Rat1 cells. We amplified 50 ng or 5 ng of poly(A)⁺ RNA by RT-PCR for 30 cycles.

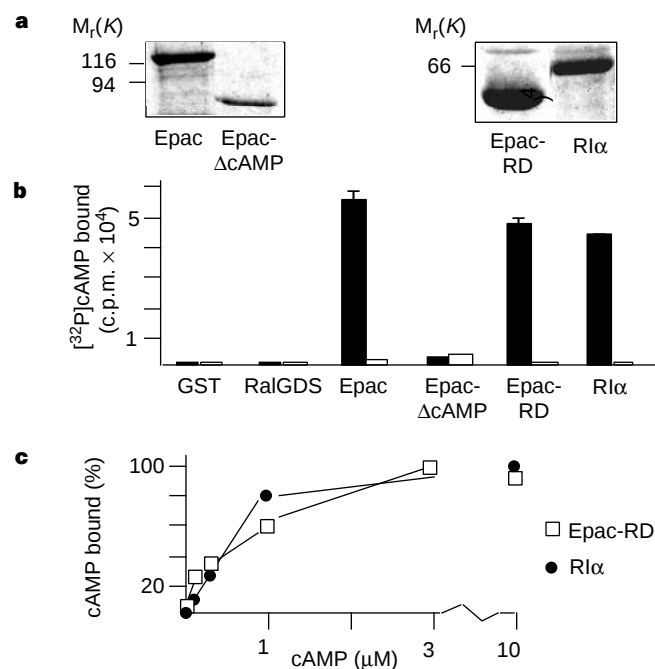


Figure 3 cAMP binds directly to the cAMP-binding domain of Epac. **a**, Coomassie blue staining of GST-fusion proteins containing Epac, Epac- Δ cAMP, Epac-RD or PKA-Rl α . **b**, We assayed 0.5 μ g GST-fusion proteins containing Epac, Epac-RD, Epac- Δ cAMP, PKA-Rl α^{25} or RalGDS-RBD for binding of $[^{32}P]cAMP$ in the absence (black bars) or presence (white bars) of excess unlabelled cAMP (means $\pm$ s.e.m., $n = 3$). **c**, Relative on/off-rate of binding of cAMP to GST-Epac-RD and GST-Rl α incubated with $[^{32}P]cAMP$ in the concentration range 3 nM to 10 μ M for 1 h.

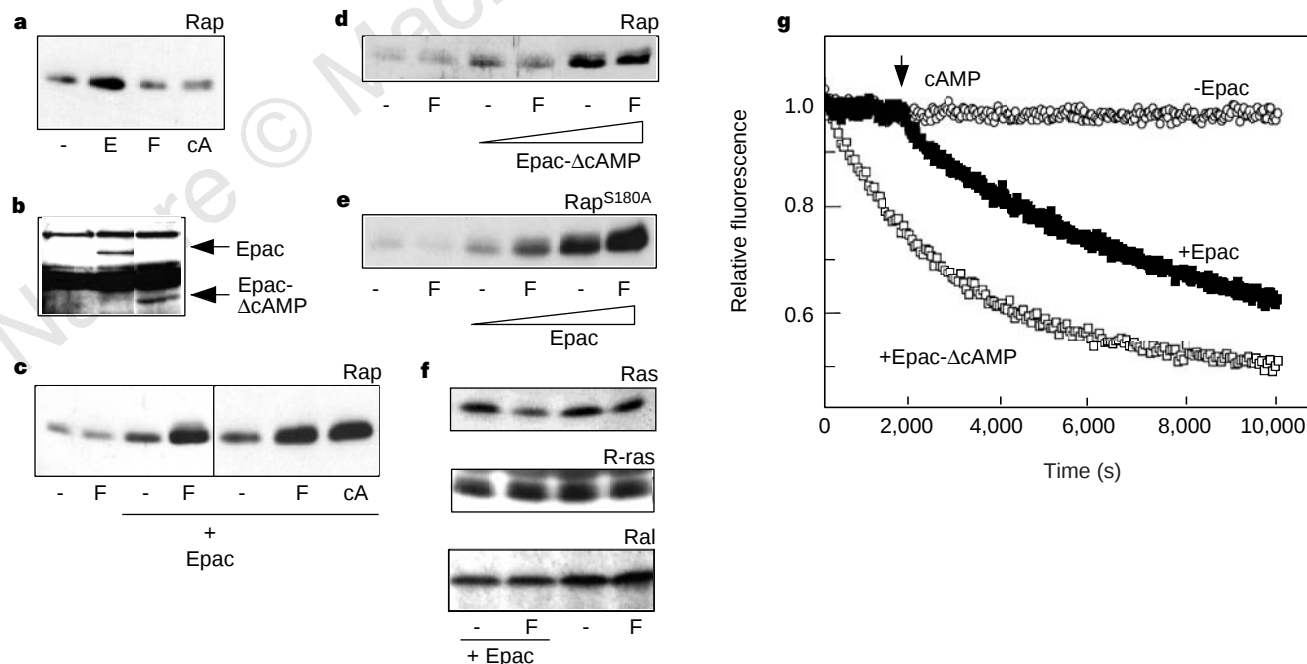


Figure 4 cAMP induces the activation of Epac in NIH3T3-A14 cells. **a**, Rap1-GTP levels in NIH3T3-A14 cells after stimulation with epidermal growth factor (EGF; E; 20 ng ml⁻¹), 8-Br-cAMP (cA; 500 μ M) or forskolin (FK; 10 μ M) for 10 min. **b**, Expression of HA-tagged Epac and Epac- Δ cAMP detected with the anti-HA monoclonal antibody 12CA5. **c**, Cells transiently expressing HA-tagged Rap1 and/or Epac were treated with forskolin (20 μ M, 10 min) or 8-Br-cAMP (1 mM, 10 min). Rap-GTP levels were analysed with 12CA5 (**c-e**). **d**, **e**, Cells expressing HA-tagged Rap1 and two concentrations of Epac- Δ cAMP or **e**, HA-tagged Rap1^{S180A} mutant and two concentrations of Epac cDNA were treated with forskolin and Rap1-GTP was analysed. **f**, Cells were transfected with the

not increased further by forskolin (Fig. 4d). The Rap1S180A mutant, which has lost the carboxy-terminal site of phosphorylation by PKA⁹, is activated by Epac and responsive to further activation by cAMP (Fig. 4e), showing that phosphorylation of Rap1 is not essential for Rap1 activation, in agreement with previous results⁹. Epac neither induced activation of Ras, R-Ras or Ral-A nor conferred responsiveness of these GTPases to cAMP (Fig. 4f). From these results we conclude that Epac is a GEF for Rap1 that is responsive to cAMP.

Finally, to confirm that Epac activates Rap1 directly and to study its regulation by cAMP, we incubated full-length Epac and Epac- Δ cAMP with Rap1A loaded with fluorescently labelled 2', 3'-bis(O)-N-methylanthranoloyl-guanosinediphosphate (mGDP) and measured the release of mGDP in real time¹⁹. Full-length Epac did not induce release of mGDP from Rap1 in the absence of cAMP, but subsequent addition of cAMP induced Epac's GEF activity (Fig. 4g). Removal of the cAMP-binding site (that is, Epac- Δ cAMP) induced *in vitro* GEF activity towards Rap1. Incubation of 20 nM Epac with equimolar or higher concentrations of cAMP resulted in full stimulation, indicating that the cAMP dissociation constant for Epac is less than 20 nM (data not shown). We conclude that cAMP regulates Epac by direct stimulation of its GEF activity and that the cAMP-binding domain of Epac acts as an inhibitory domain that affects GEF activity either allosterically or by steric hindrance.

Epac represents the first step of a new cAMP-induced signalling pathway, which suggests that not all cAMP-induced effects are mediated by either PKA or cyclic-nucleotide-gated channels, the only previously known cAMP-target proteins. Several reports have suggested the existence of such pathways (for example, that leading to DNA replication in thyroid cells²⁰). Epac activates Rap1, a ubiquitously expressed small GTPase whose precise function is

indicated GTPases (HA-tagged) with or without Epac, and were or were not stimulated with forskolin; the GTP-bound forms of the GTPases were identified. **g**, Rap1A was loaded with fluorescent mGDP and mGDP release was measured in real time in the absence (open circles) or presence of 50 nM Epac (filled squares) or Epac- Δ cAMP (open squares). After 30 min, 10 μ M cAMP was added to the incubation mixture containing Epac. cAMP did not affect intrinsic exchange of guanine nucleotides on Rap1 or Epac- Δ cAMP-induced guanine-nucleotide exchange on Rap1. Also, Epac- Δ cAMP failed to induce mGDP release from Ras (data not shown).

still unknown. Rap1 may antagonize Ras-mediated signalling²¹ and support signalling by Ras effectors¹⁰. However, Rap1 may also have a Ras-independent function⁷. Activation of Rap1 is a common event after the stimulation of several cell types by extracellular stimuli, and, depending on the cell type, Ca²⁺, diacylglycerol and cAMP mediate this activation^{3,5,6}. Our discovery of a cAMP-responsive GEF for Rap1 reveals the mechanism by which one of these activation pathways operates. Interestingly, a serine close to the C terminus of Rap1 is phosphorylated by PKA⁹, showing that cAMP may affect Rap1 signalling in a PKA-dependent manner as well. □

Methods

Cells. Rat1 and NIH3T3-A14 cells were grown in DMEM with 10% fetal calf serum (FCS) and CHO10001 and CHO10248 cells were grown in DMEM with 7.5% FCS. Rat1 and NIH3T3-A14 cells were then starved for 16 h in DMEM with 0.5% FCS, and CHO cells were starved for 60 min in 10 mM HEPES, pH 7.4, 135 mM NaCl, 5.4 mM KCl, 1.4 mM CaCl₂, 1.4 mM MgSO₄, 0.2 mM sodium phosphate, 5 mM glucose, 0.1% bovine serum albumin (BSA).

Rap1-activation assay. Cells were lysed in Ral buffer²² and 1 ml cleared lysate was incubated with 5 µg GST-RalGDS-RBD bound to glutathione beads. After rotation for 45 min at 4 °C, beads were washed four times and boiled in sample buffer. Samples were separated by 15% SDS-PAGE, blotted to poly(vinylidene fluoride) (PVDF) membrane, probed with anti-Rap1 antibody (Transduction Laboratories) or anti-haemagglutinin (HA) monoclonal antibody 12CA5 (Fig. 4) and visualized by enhanced chemiluminescence (Amersham)⁶. The GTP-bound forms of Ras, R-Ras and Ral-A were recovered using GST-Raf1RBD (for Ras and R-Ras) and GST-RalBD of RLIP76 (for Ral-A)²².

Isolation of Epac cDNA. In a database search we identified a genomic sequence encoding a protein with a putative cAMP-binding site and a putative GEF domain (*Homo sapiens* PAC RPC13-197B17, Genebank accession number AC004241). PCR primers with flanking restriction sites (lower-case letters) were selected on the basis of intron/exon predictions for the genomic sequence by NIX analysis (<http://www.hgmp.mrc.ac.uk/>); 5' primer: 5'-gtcgacGGTGTGAGAAGGATGAC-3'; 3' primer: 5'-gcggccgCATGGCTCCAGCTCTCG-3'. Epac cDNA was retrieved from a human muscle cDNA library and cloned into the *SalI*-*NotI* sites of the pMT2SM-HA expression vector. The sequence was determined by dideoxy-sequencing.

Epac-ΔcAMP was constructed by PCR using a 5' primer spanning amino acids 322–329, 5'-gtcgacaCTGGAGAGAGCCTCTCAG-3', and the 3' primer mentioned above. To study *in vitro* binding of cAMP, cDNAs encoding Epac and Epac-ΔcAMP were cloned into the *XhoI*-*NotI* sites of the pGEX 4T3 vector. Epac-RD was cloned by PCR using the 5' primer and a 3' primer spanning amino acids 329–322, 5'-gcggccgCTGAGAGGCTCTCTCCAG-3', in the same vector. Sequences of all inserts were confirmed.

GST-fusion constructs were introduced into BL-21 bacteria and protein expression was induced with 0.1 µM isopropyl-β-D-thiogalactoside (IPTG) at room temperature for 20 h. Bacteria were lysed in PBS containing 1% Triton-X100, 1 mM dithiothreitol (DTT), 1 µM aprotinin and 1 µM leupeptin. Lysates were sonicated and insoluble material was removed by centrifugation. Cleared lysates were incubated with 100 µl glutathione-agarose beads for 30 min. Beads were washed four times in PBS and 1 mM DTT and analysed by SDS-PAGE. For the *in vitro* exchange assays, GST-fusion proteins were eluted from the beads in PBS with 50 mM TrisCl, pH 7.4, 1 mM DTT, 10 mM glutathione and 10% glycerol. Proteins were dialysed in 50 mM TrisCl, pH 7.4, 150 mM NaCl and 10% glycerol.

Expression of Epac. Expression of Epac was measured on a human multi-tissue northern blot (Clontech) using ³²P-labelled Epac cDNA as a probe. The presence of Epac mRNA in NIH3T3-A14 and Rat1 cells was determined by RT-PCR analysis using the following primers: 5'-CTTCTCCAGAACTCTCAG-3' and 5'-TCAGCTCATGCACTTCTCTG-3'. These primers were designed from a region that is identical in mouse and human Epac. The PCR products were blotted and probed with ³²P-labelled Epac cDNA to confirm their identity.

cAMP-binding assay. Assays of [³²P]cAMP (ICN) binding were done by modification of the method of ref. 23 for cAMP binding by PKA. We incubated 0.5 µg GST-fusion protein bound to glutathione beads in 100 µl cAMP-binding mix (10 mM potassium phosphate, pH 6.8, 2 M NaCl, 1 mM EDTA,

10 µg BSA and 15 nM [³²P]cAMP) with or without 10 mM unlabelled cAMP for 45 min at room temperature. To determine binding of [³²P]cAMP to the fusion protein, we rapidly washed the beads four times with ice-cold 10 mM potassium phosphate and 1 mM EDTA, pH 6.8, and measured bound [³²P]cAMP using a scintillation counter. To compare the affinity of cAMP for Epac and the RIα subunit of PKA, we incubated Epac-RD and RIα with different concentrations of [³²P]cAMP for 1 h at room temperature and treated them as above.

In vitro Rap1 activation. We loaded 200 nM Rap1A with fluorescent mGFP as described¹⁹ and incubated the labelled Rap1A with ~50 nM Epac in 50 mM Tris, pH 7.6, 5 mM MgCl₂, and 2 mM dithioerythrol at 20 °C. The reaction was started by addition of excess (40 µM) unlabelled GDP. When indicated, 50 µM 8-Br-cAMP was added. Release of mGDP results in a decrease in fluorescence, which we measured in a Perkin Elmer LS50 spectrometer as described¹⁹. To determine the concentration of cAMP that induces half-maximal activity of Epac *in vitro*, we incubated 20 nM Epac with different concentrations of 8-Br-cAMP.

Received 16 September; accepted 6 November 1998.

- Kitayama, H., Sugimoto, Y., Matsuzaki, T., Ikawa, Y. & Noda, M. A ras-related gene with transformation suppressor activity. *Cell* **56**, 77–84 (1989).
- Boussiotis, V. A., Freeman, G. J., Berezovskaya, A., Barber, D. L. & Nadler, L. M. Maintenance of human T cell anergy: blocking of IL-2 gene transcription by activated Rap1. *Science* **278**, 124–128 (1997).
- Reedquist, K. A. & Bos, J. L. Costimulation through CD28 suppresses T cell receptor-dependent activation of the Ras-like small GTPase Rap1 in human T lymphocytes. *J. Biol. Chem.* **273**, 4944–4949 (1998).
- York, R. D. *et al.* Rap1 mediates sustained MAP kinase activation induced by nerve growth factor. *Nature* **392**, 622–626 (1998).
- M'Rabet, L. *et al.* Activation of Rap1 in human neutrophils. *Blood* **92**, 2133–2140 (1998).
- Zwartkruis, F. J. T., Wolthuis, R. M. F., Nabben, N. M. J. M., Franke, B. & Bos, J. L. Extracellular signal-regulated activation of Rap1 fails to interfere in Ras effector signalling. *EMBO J.* **17**, 5905–5912 (1998).
- Franke, B., Akkerman, J. W. & Bos, J. L. Rapid Ca²⁺-mediated activation of Rap1 in human platelets. *EMBO J.* **16**, 252–259 (1997).
- Altschuler, D. L. & Ribeiro-Neto, F. Mitogenic and oncogenic properties of the small G protein Rap1b. *Proc. Natl Acad. Sci. USA* **95**, 7475–7479 (1998).
- Altschuler, D. L., Peterson, S. N., Ostrowski, M. C. & Lapetina, E. G. Cyclic AMP-dependent activation of Rap1b. *J. Biol. Chem.* **270**, 10373–10376 (1995).
- Vossler, M. R. *et al.* cAMP activates MAP kinase and Elk 1 through a Raf- and Rap1 dependent pathway. *Cell* **89**, 74–82 (1997).
- Singh, T. J. *et al.* Characterization of a cyclic AMP-resistant Chinese hamster ovary cell mutant containing both wild-type and mutant species of type I regulatory subunit of cyclic AMP-dependent protein kinase. *J. Biol. Chem.* **260**, 13927–13933 (1985).
- Ebinu, J. O. *et al.* RasGRP, a Ras guanyl nucleotide-releasing protein with calcium- and diacylglycerol-binding motifs. *Science* **280**, 1082–1086 (1998).
- Zagotta, W. N. & Siegelbaum, S. A. Structure and function of cyclic nucleotide-gated channels. *Annu. Rev. Neurosci.* **19**, 235–263 (1996).
- Shabb, J. B., Ng, L. & Corbin, J. D. One amino acid change produces a high affinity cGMP-binding site in cAMP-dependent protein kinase. *J. Biol. Chem.* **265**, 16031–16034 (1990).
- Gotoh, T. *et al.* Identification of Rap1 as a target for the Crk SH3 domain-binding guanine nucleotide-releasing factor C3G. *Mol. Cell. Biol.* **15**, 6746–6753 (1995).
- Boriack-Sjodin, P. A., Margarit, S. M., Bar-Sagi, D. & Kuriyan, J. The structural basis of the activation of Ras by Sos. *Nature* **394**, 337–343 (1998).
- Ponting, C. P. & Bork, P. Pleckstrins repeat performance: a novel domain in G-protein signaling? *Trends Biochem. Sci.* **21**, 245–256 (1996).
- Axelrod, J. D., Miller, J. R., Shulman, J. M., Moon, R. T. & Perrimon, N. Differential recruitment of dishevelled provides signaling specificity in the planar cell polarity and wingless signaling pathways. *Genes Dev.* **12**, 2610–2622 (1998).
- van den Berge, N., Cool, R. H., Horn, G. & Wittinghofer, A. Biochemical characterization of C3G: an exchange factor that discriminates between Rap1 and Rap2 and is not inhibited by Rap1A(S17N). *Oncogene* **15**, 845–850 (1997).
- Dremier, S. *et al.* Activation of cyclic AMP-dependent kinase is required but may not be sufficient to mimic cyclic AMP-dependent DNA synthesis and thyroglobulin expression in dog thyroid cells. *Mol. Cell. Biol.* **17**, 6717–6726 (1997).
- Cook, S. J., Rubinfeld, B., Albert, I. & McCormick, F. RapV12 antagonizes Ras-dependent activation of ERK1 and ERK2 by LPA and EGF in Rat-1 fibroblasts. *EMBO J.* **12**, 3475–3485 (1993).
- Wolthuis, R. M. *et al.* Activation of the small GTPase Ral in platelets. *Mol. Cell. Biol.* **18**, 2486–2491 (1998).
- Rannels, S. R. & Corbin, J. D. Using analogs to study selectivity and cooperativity of cyclin nucleotide binding sites. *Methods Enzymol.* **99**, 168–175 (1983).
- Verheijen, M. G. H. & Defize, L. H. K. Parathyroid hormone inhibits mitogen-activated protein kinase activation in osteosarcoma cells via a protein kinase A-dependent pathway. *Endocrinology* **136**, 3331–3337 (1995).
- Herberg, F. W., Zimmermann, B., McGlone, M. & Taylor, S. S. Importance of the A-helix of the catalytic subunit of cAMP-dependent protein kinase for stability and for orienting subdomains at the cleft interface. *Protein Sci.* **6**, 569–579 (1997).

Acknowledgements. We thank F. W. Herberg for the GST-RIα subunit of PKA, M. Gottesman for CHO10001/10248 cells, B. M. T. Burgering for discussions and for reading the manuscript and the rest of our colleagues for continuous support and discussions. J.deR. and M.H.G.V. were supported by the Council of Earth and Life Sciences of the Netherlands Organization for Scientific Research, F.J.T.Z. by the Dutch Cancer Society and R.H.C. by the EG.

Correspondence and requests for materials should be addressed to J.L.B. (e-mail: j.l.bos@med.uu.nl). The Epac gene sequence has been deposited with Genebank under accession number AF 103905.

Functional interaction between InsP_3 receptors and store-operated Htrp3 channels

Kirill Kiselyov*, Xin Xu*, Galina Mozhayeva†, Tuan Kuo‡, Isaac Pessah§, Gregory Mignery||, Xi Zhu¶, Lutz Birnbaumer# & Shmuel Muallem*

* Department of Physiology, University of Texas Southwestern Medical Center at Dallas, 5323 Harry Hines Blvd, Dallas, Texas 75235, USA

† Institute of Cytology, St Petersburg 194064, Russia

‡ Wayne State University, Detroit, Michigan 48201, USA

§ Department of Molecular Biosciences, School of Veterinary Medicine, University of California, Davis, California 95616, USA

|| Department of Physiology, Stritch School of Medicine, Loyola University Chicago, Maywood, Illinois 60153, USA

¶ Department of Pharmacology and Neurobiotechnology Center, The Ohio State University, Columbus, Ohio 43016, USA

Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles, California 90049, USA

Calcium ions are released from intracellular stores in response to agonist-stimulated production of inositol 1,4,5-trisphosphate (InsP_3), a second messenger generated at the cell membrane. Depletion of Ca^{2+} from internal stores triggers a capacitative influx of extracellular Ca^{2+} across the plasma membrane^{1,2}. The influx of Ca^{2+} can be recorded as store-operated channels (SOC) in the plasma membrane or as a current known as the Ca^{2+} -release-activated current (I_{crac})^{3–5}. A critical question in cell signalling is how SOC and I_{crac} sense and respond to Ca^{2+} -store depletion: in one model, a messenger molecule is generated that activates Ca^{2+} entry in response to store depletion^{1,6}; in an alternative model⁷, InsP_3 receptors in the stores are coupled to SOC and I_{crac} . The mammalian Htrp3 protein⁸ forms a well defined store-operated channel^{8,9} and so provides a suitable system for studying the effect of Ca^{2+} -store depletion on SOC and I_{crac} . We show here that Htrp3 channels stably expressed in HEK293 cells are in a tight functional interaction with the InsP_3 receptors. Htrp3 channels present in the same plasma membrane patch can be activated by Ca^{2+} mobilization in intact cells and by InsP_3 in excised patches. This activation of Htrp3 by InsP_3 is lost on extensive washing of excised patches but is restored by addition of native or recombinant InsP_3 -bound InsP_3 receptors. Our results provide evidence for the coupling hypothesis⁷, in which InsP_3 receptors activated by InsP_3 interact with SOC and regulate I_{crac} .

A typical response of control HEK293 cells to application of Ca^{2+} -mobilizing agonists is the activation of a miniature Ca^{2+} channel (Fig. 1a). As the properties of this channel are similar to those described for miniature channels in several cell types^{10,11}, it may be the native store-operated channel of HEK293 cells. Under the same conditions, the majority (124 out of 155) of Htrp3-expressing cells (T3 cells) respond by activation of a large, non-selective channel (Fig. 1b, c) that was seen only in T3 cells (Fig. 2a) and had the same properties as Htrp3. In agreement with earlier results^{9,12,13}, the channel had a major conductance of 66 pS (Fig. 2b), but from the previous results^{9,12,13} and those shown in Fig. 1, together with all-point amplitude histograms in Fig. 2f and noise analysis (data not shown), Htrp3 also has a 17-pS conductance (Fig. 2b, d–f). Both states have similar kinetic parameters (Fig. 2) and were almost equally prevalent in most records (Fig. 2d). Htrp3 had the same permeability for K^+ , Na^+ , Ca^{2+} and Ba^{2+} , and its open probability increased at membrane potentials above +40 mV (Fig. 2c). The 17-pS channels may assemble into a 66-pS complex, or the 66-pS channel may have a 17-pS substate; these possibilities are not investigated here.

A central finding is that Htrp3 can be activated in isolated, cell-free

plasma membrane patches by application of InsP_3 to the cytosolic face of a patch. Excision of the membrane patch containing active Htrp3 caused a loss of channel activity, which could then be restored by 2 μM InsP_3 (Fig. 1b). The kinetic properties of channels activated by agonist in intact cells or by InsP_3 in excised patches were the same, including conductance, ion selectivity, voltage-dependence of the open probability (NP_o) and mean open time (Fig. 2). Furthermore, an inhibitor of SOC/ I_{crac} (ref. 14) and Htrp3 channels¹⁵, SKF 96265 (SKF), inhibited the channels activated by InsP_3 in excised patches (Fig. 3a). The effect of InsP_3 on Htrp3 was inhibited by heparin (Fig. 3b) and by the membrane-permeable inhibitor of the InsP_3 receptor (InsP_3R), xestospongine C (Xest C; Fig. 3c)¹⁶. Considering that InsP_3 was the only second messenger present under excised-patch conditions, the results shown in Figs 1–3 indicate that InsP_3 receptors communicate with Htrp3 and regulate its activity, in agreement with the coupling hypothesis.

The dependence of Htrp3 activity on the Ca^{2+} content of internal stores was tested by using thapsigargin (Tg), which depletes stored Ca^{2+} at the InsP_3 concentrations present in resting cells^{17,18}. Thapsigargin-induced store depletion in intact cells and cytosolic InsP_3 in

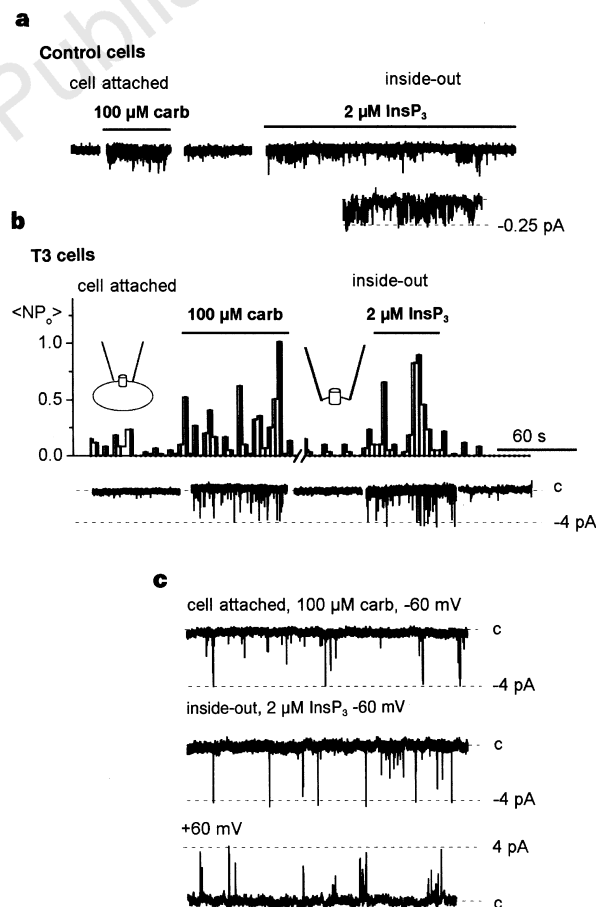


Figure 1 Activation of Htrp3 channels by Ca^{2+} -mobilizing agonists and InsP_3 . The cell-attached mode of the patch-clamp technique was used to monitor channel activity in plasma membrane patches of intact control HEK293 and Htrp3-expressing cells (T3 cells) and in membrane patches excised from them. **a**, Ca^{2+} current in the control cell was recorded at a holding potential of -60 mV. As indicated, the cell was stimulated with 100 μM carbachol (carb) and subsequently the patch was excised and exposed to 2–10 μM InsP_3 . Bottom right, a segment of the InsP_3 -activated current on an expanded timescale. **b**, The same protocol was used to record the activity of Htrp3 in T3 cells. The models indicate the experimental protocol. **c**, Htrp3 channel activity at an expanded timescale; the letter 'c' denotes the closed state, and the dashed lines represent channel openings to 66-pS state.

excised patches activated Htrp3 present in the same plasma membrane patch (Fig. 4a, b). Thapsigargin was not as effective as agonists in stimulating Htrp3, suggesting that both depleted stores and InsP_3 -bound InsP_3 receptors are required for the activation of Htrp3 (see below). In agreement with this suggestion, we found that stimulation of thapsigargin-treated cells with agonist, which increases the cytosolic InsP_3 concentration after store depletion, further stimulated Htrp3 (Fig. 4c). These results are equivalent to the increased Ca^{2+} and Ba^{2+} influx that occurs in thapsigargin-treated T3 cells stimulated with carbachol¹⁵.

The relation between the activities of the InsP_3 receptor and Htrp3 is confirmed by the requirement of an active InsP_3 receptor for activation of Htrp3. Figure 5a shows that irreversible depletion of internal stores by treatment of T3 cells with agonist and thapsigargin caused a persistent activation of Htrp3. Inhibition of channel activity by addition of 20 μM SKF (Fig. 5a, c) to the bath indicated that the current was carried by Htrp3. Diffusion of heparin into the cytosol

after Htrp3 activation effectively inhibited Htrp3 activity, despite the continued presence of thapsigargin which maintains the stores in a depleted state (Fig. 5b). In the cell-attached mode, Htrp3 was activated by thapsigargin in the absence (Fig. 5d, f) or presence (Fig. 5e, f) of agonist. Incubation of these cells with Xest C markedly inhibited Htrp3 activity. These results indicate that an active InsP_3 receptor is required for maintaining Htrp3 in the active state.

The functional coupling between InsP_3 receptors and Htrp3 suggests that stable expression of Htrp3 should be reflected in a concomitant change in the number of InsP_3 receptors. Figure 6a shows that expression of Htrp3 increased the expression of all InsP_3 receptor isoforms in HEK293 cells. These findings can explain why cells transiently transfected with large amounts of Htrp3 (~600 channels per cell) show high spontaneous activity¹³, whereas in T3 cells (~20 Htrp3 channels per cell) the spontaneous activity of Htrp3 is low. It is likely that T3 cells adjust to the stable expression of Htrp3 by upregulating their expression of InsP_3 receptors.

The Htrp3 and InsP_3 receptors did not coimmunoprecipitate and neither could InsP_3 activate Htrp3 if the patches were extensively washed before application of InsP_3 (Fig. 6b). This suggests that there was a weak interaction between the channels. But we were able to reconstitute the Htrp3– InsP_3R complex by adding native or recombinant InsP_3 receptors to the patches. We used rat or bovine cerebellar microsomes as a source of native InsP_3 receptors¹⁹, and rat forebrain microsomes, which are a poor source of InsP_3 receptors¹⁹, as a control. Addition of cerebellar microsomes to the solution containing InsP_3 strongly activated Htrp3, whereas omission of InsP_3 or microsomes caused channel activity to drop to zero (Fig. 6). We were able to restore Htrp3 activity in ~25% of experiments, suggesting that targeting of the Htrp3 channel by the InsP_3 receptor is a stochastic event and depends on how a microsome binds to the patch. Persistent Htrp3 reactivation occurred in 3/12 experiments (Fig. 6b), and in 3/12 experiments the microsomes only transiently restored Htrp3 activity (Fig. 6d), as though binding was unstable. In 6/12 attempts, Htrp3 was not

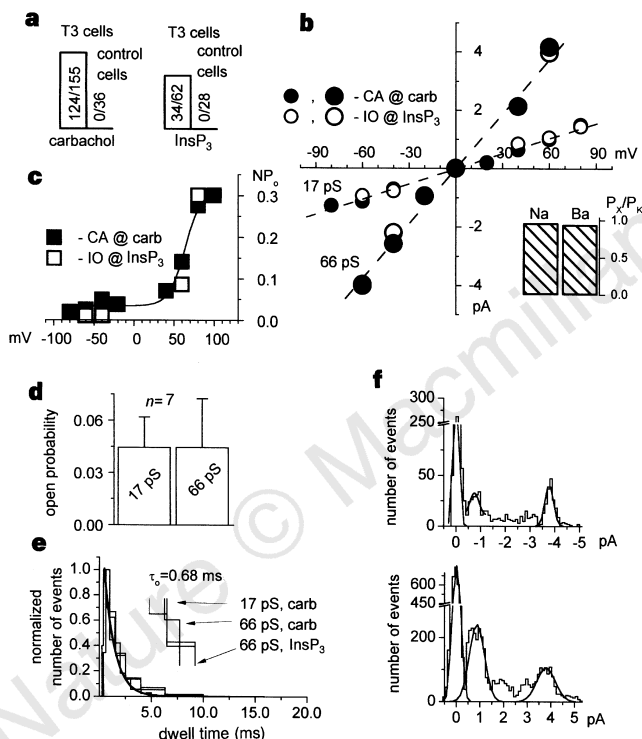


Figure 2 Kinetic properties of the channels activated by agonists in intact cells and by InsP_3 in excised patches. **a**, Summary of Htrp3 occurrence in T3 and control cells stimulated with carbachol in cell-attached and with InsP_3 in inside-out mode. **b**, The current-voltage relation for agonist (filled circles) and InsP_3 -activated (open circles) channels (each point represents data from 2–7 experiments); large symbols, 66-pS conductance; small symbols, 17-pS conductance. Error bars (s.e.m.) are smaller than symbol size. Inset, relative permeabilities (P_x , calculated from reversal potentials) of the channels for Ba^{2+} , Na^+ and K^+ . I - V curves for Ba^{2+} (not shown) were derived from 4 experiments. CA@carb, cell-attached and carbachol; IO@ InsP_3 , inside-out and InsP_3 ; P_K , K^+ permeability. **c**, A plot of Htrp3 NP_0 activated by agonist (filled squares) or InsP_3 (open squares) as a function of membrane potential. **d**, Probability of Htrp3 channel being open to 17-pS and 66-pS states. Data were obtained from 100–300-s current recordings. **e**, The dwell-time distributions for the 66-pS conductance activated by carbachol or InsP_3 and the 17-pS conductance activated by carbachol are superimposed. Data were normalized to facilitate comparison between the different conditions. The average dwell times were calculated from single exponential fits (thick line). The line represents fits for all three curves. **f**, All-point amplitude histograms were obtained at -60 (top) and +60 (bottom) mV from a representative record of Htrp3 activity. Note the presence of the two peaks corresponding to 17 and 66 pS. Peaks were fitted by gaussian distributions.

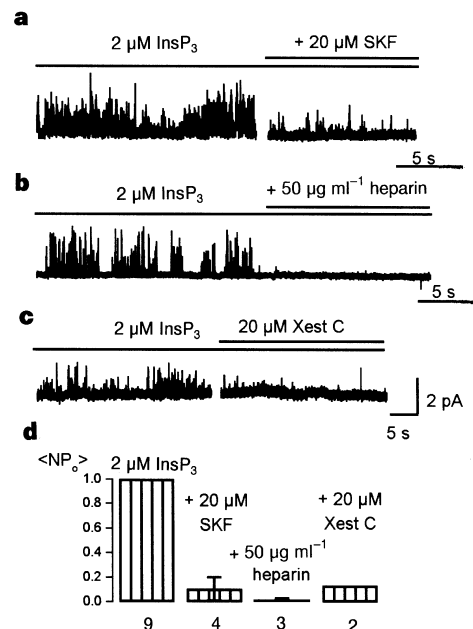


Figure 3 The effect of Htrp3 and InsP_3R channel inhibitors on InsP_3 -induced activation. The experiments were done with inside-out patches. In all experiments, I - V relations were determined before and after application of the indicated inhibitor to identify the channels. **a**, SKF and **b**, heparin were added to the bath solution; **c**, Xest C was applied by puffing from an adjacent pipette. **d**, Summary of results from the number of experiments shown below the graph. NP_0 was averaged from 30–60 s current recordings before and after application of inhibitors.

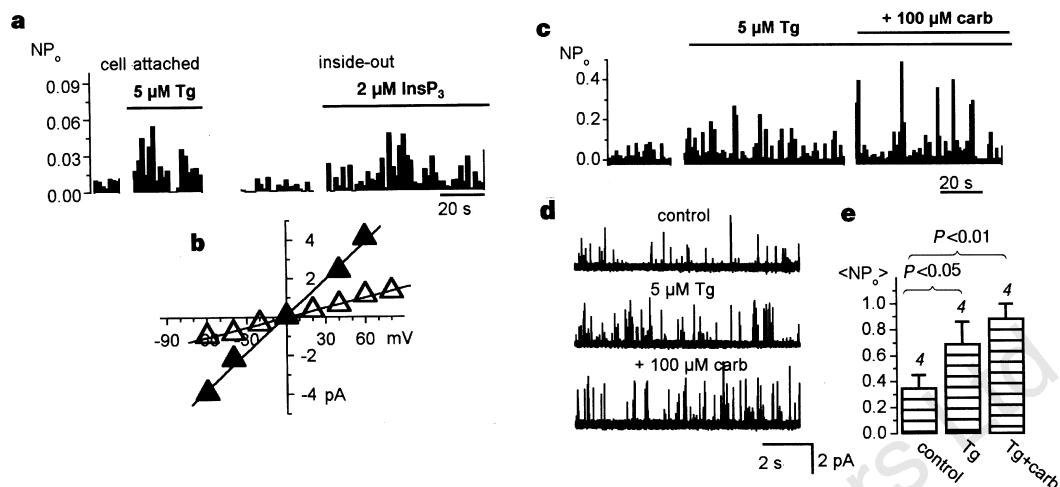


Figure 4 Activation of Htrp3 channels by thapsigargin and InsP₃. **a**, The protocol used for Fig. 1 was also used to study Htrp3 activation by Tg and InsP₃. **b**, I-V curve for Tg-induced channel activity. Each point represents the results of 2-6 experiments. Error bars (s.e.m.) are smaller than the symbol size. Slopes equal

16.8 pS (triangles) and 65 pS (filled triangles). **c**, NP₀ of Htrp3 induced by Tg and subsequent stimulation of the same cell by carbachol. **d**, Expanded segments of the current recording shown in **c**. **e**, Summary of results of the 4 experiments.

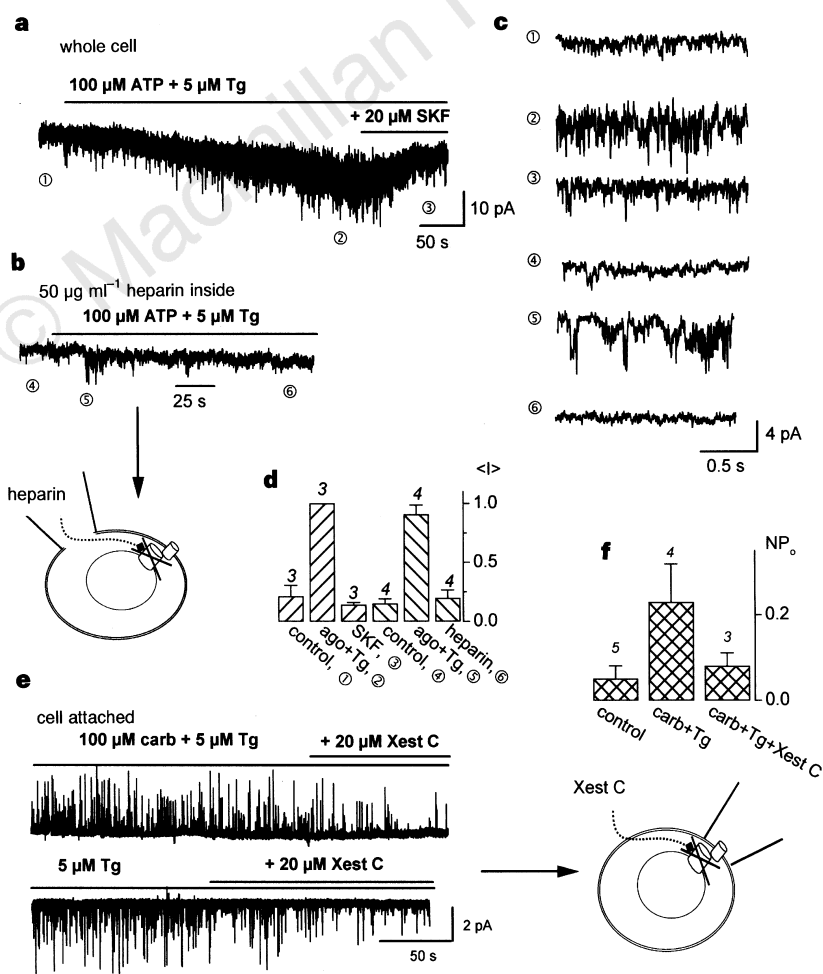


Figure 5 Suppression of Htrp3 activity by inhibitors of InsP₃R in intact cells. **a-d**, The whole-cell mode of the patch-clamp technique was used to monitor channel activity. Htrp3 channels were activated by agonist and Tg in the **a**, absence, or **b**, presence of heparin in the pipette solution. In **a**, the indicated channels were inhibited by SKF. **c**, Segments of channel recordings (numbered) from the

experiments in **a** and **b** at an expanded timescale. Experiments are summarized in **d**. Conditions and periods of estimation of channel activity are listed below the bars. In **e** and **f**, Htrp3 activity was monitored in cell-attached mode. Where indicated in **e**, cells treated with Tg in the presence (top) and absence (bottom) of carbachol were exposed to Xest C.

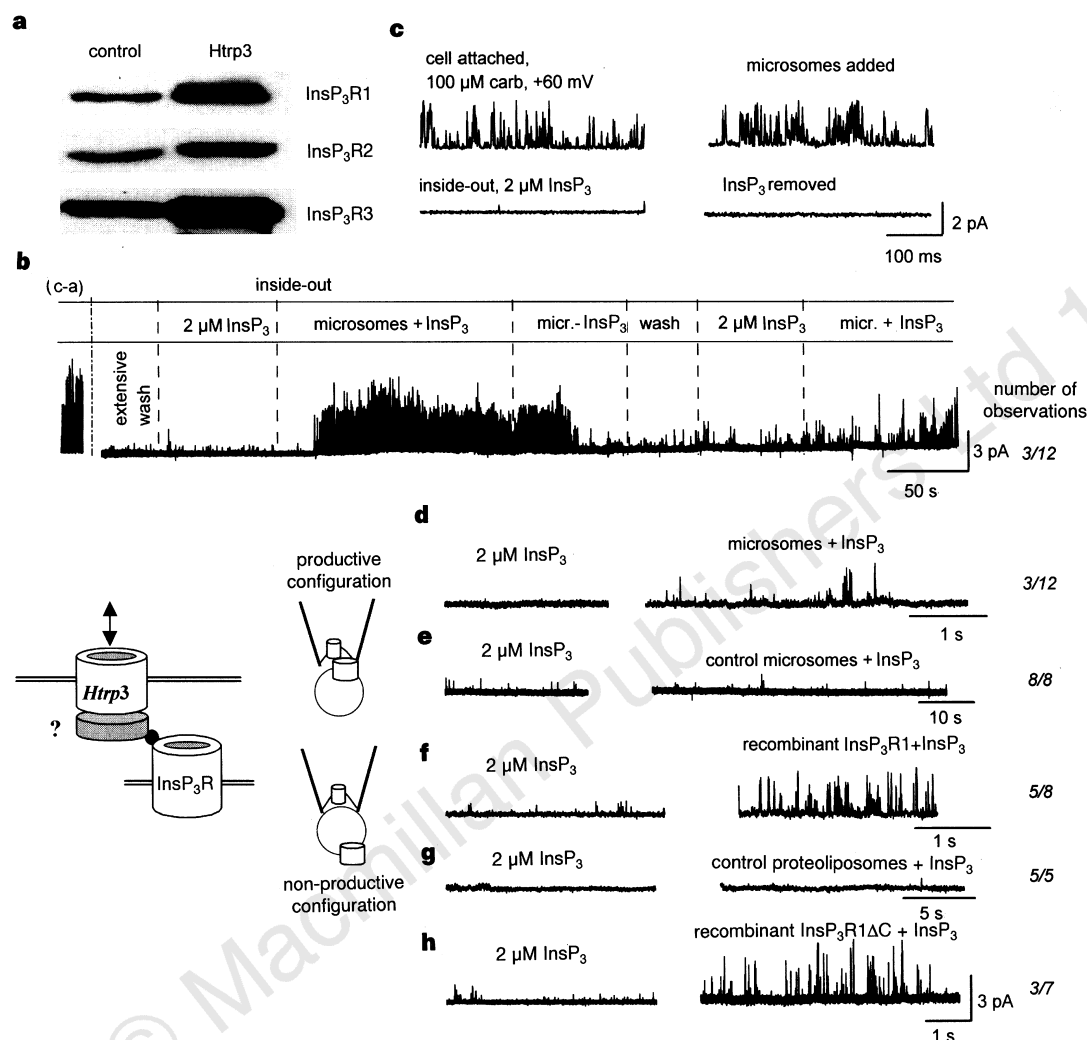


Figure 6 Reconstruction of Htrp3-InsP₃ receptor complexes. **a**, Western blot analysis of InsP₃R expression in control and T3 cells, performed as before²⁶. **b-h**, The protocol used for Fig. 1 was used to find the effect of **b-d**, cerebellar and **f-h**, recombinant InsP₃R on Htrp3 activity. Initially, the currents were recorded in the cell-attached (**c-a**) mode to ensure that the patch contained the Htrp3 channel. The channel was then completely inactivated by patch excision and extensive washes until application of InsP₃ evoked no channel openings. The addition of bovine cerebellar microsomes (**b-d**) caused robust and stable (**b, c**) or transient

(**d**) activation of Htrp3 channel. Removal of InsP₃ or microsomes inactivated the Htrp3 channels (**b, c**). The numbers of successful attempts are listed in the right column. **c**, Segments of current recordings at an expanded timescale from those in **b, c, g**. Typical experiments with control microsomes. **f, h**, Examples of activation of Htrp3 by recombinant InsP₃R reconstituted into liposomes (**f**) and microsomes prepared from HEK293 cells expressing a C-terminally truncated ΔCInsP₃R1 (**h**). Models on the left illustrate the proposed interactions under our experimental conditions.

activated, possibly because of impaired access of the InsP₃ receptor to a patch capped by a microsome free of InsP₃ receptors (see model in Fig. 6). Recombinant InsP₃R1, which had been partially purified on a sucrose gradient and reconstituted into liposomes, could also activate the Htrp3 channel when added together with InsP₃ to inside-out patches (5/8 attempts; Fig. 6f). Neither control forebrain microsomes (8/8 attempts) nor control proteoliposomes (5/5 attempts) activated Htrp3 channels (Fig. 6e, g).

The activation of Htrp3 by InsP₃R1 appears to contradict an earlier report²⁰ that deletion of all known isoforms of the InsP₃ receptor does not prevent stimulation of Ca²⁺ influx by thapsigargin. However, in these experiments²⁰ native InsP₃ receptors were replaced with truncated constructs: for example, only 100 amino acids of the C terminus of InsP₃R1 were deleted. Although this inactivated InsP₃R1 channel activity, it may have not prevented regulation of Ca²⁺ influx by InsP₃R1. To test this idea, we removed 154 amino acids from the C terminus of InsP₃R1 (ΔCInsP₃R1) and tested its activity. Figure 6h shows that ΔCInsP₃R1 was as effective as full-length InsP₃R1 in activating Htrp3. Furthermore, it is likely that

the cells, which were triple mutants for InsP₃ receptors²⁰, expressed other Ca²⁺-releasing channels, such as ryanodine receptors, which mediated the activation of Ca²⁺ influx by thapsigargin.

We have used molecular and functional approaches to show that activation of Htrp3 channels requires both Ca²⁺ depletion from internal stores and InsP₃ receptors occupied with InsP₃. The stimulation of Htrp3 activity by agonist in thapsigargin-treated cells and its inhibition by inhibitors of InsP₃ receptors in intact cells indicate that, in addition to Ca²⁺-store depletion, the binding of InsP₃ to its receptor is required for Htrp3 activation; this conclusion is supported by the activation by InsP₃ of Htrp3 in excised plasma membrane patches. The loss of this activation in extensively washed patches and its restoration after incubating the patches with native or recombinant InsP₃ receptor suggests that the two channels interact either directly or indirectly. The regulation of Htrp3 by stored Ca²⁺ by a mechanism similar to that used to regulate *I*_{crac}, together with a recent demonstration of an *I*_{crac}-like current in excised patches that is activated by InsP₃ (refs 10, 11), suggests that the interaction between plasma membrane channels and InsP₃ receptors is used generally for regulating

store-dependent Ca^{2+} influx. Such regulation is reminiscent of excitation–contraction coupling in skeletal muscle^{21,22}. Like the preferred localization of ryanodine receptors in terminal cisternae²², microdomains of highly expressed InsP_3 receptors next to the plasma membrane²³ may form a site of regulatory interaction between the InsP_3 receptors and Ca^{2+} -influx channels. Irrespective of the nature of the interaction between the channels, our findings strongly support the coupling hypothesis^{2,7}. □

Methods

Cells. Control cells and clone 56 of HEK293 cells stably expressing Htrp3 (ref. 8) were grown in DMEM medium supplemented with 10% FBS, 1% penicillin/streptomycin and $200 \mu\text{g ml}^{-1}$ G418 (Gibco BRL) and maintained at 37°C in 5% CO_2 . One day before the experiments, cells were transferred to glass coverslips for patch-clamp recordings.

Electrophysiology. In the cell-attached mode, the pipette solution contained 100 mM Na/HEPES and 2 mM CaCl_2 ; pH of all solutions was 7.4. Before changing to high- K^+ solution, cells were bathed in normal Ringer solution containing (in mM): 140 NaCl, 10 KCl, 1 MgCl_2 , 2 CaCl_2 , 10 HEPES. After a seal was established, the medium was replaced with a high- K^+ solution containing (in mM): 145 KCl, 5 NaCl, 1 MgCl_2 , 2 CaCl_2 , 10 HEPES. For estimation of channel permeability for Ba^{2+} , pipettes were filled with 100 mM BaCl_2 , 10 HEPES, or 100 mM Ba^{2+} /HEPES, 2 BaCl_2 . Patches were excised into a solution containing (in mM): 140 potassium gluconate, 10 HEPES, 2 EGTA, 1 MgCl_2 , 5 NaCl, 1.13 CaCl_2 (pCa7), or a solution without CaCl_2 . For whole-cell studies, the pipette solution contained (in mM): 120 potassium gluconate, 20 HEPES, 10 EGTA, 1 MgCl_2 , 5 NaCl, 5 ATP. Bath solution during whole-cell experiments was normal Ringer. Current was recorded with an Axopatch 200B patch-clamp amplifier equipped with a Digidata 1200 PC interface. Currents were recorded at 2–5 kHz, filtered at 0.2–1 kHz and analysed with Axoscope 1.1, PClamp 6.0.3 (Axon Instruments) and Origin 5.0 (Microcal) software. NP_0 was determined using the following equation: $\text{NP}_0 = \langle I \rangle / i$, where $\langle I \rangle$ and i are the mean channel current and unitary current amplitude, respectively.

Cerebellar microsomes. Bovine or rat cerebella (or forebrains) were homogenized in a solution containing (in mM): 50 Tris, 100 NaCl and 2 EDTA. The homogenate was centrifuged for 10 min at 500g and the supernatant was collected and centrifuged for 25 min at 40,000g. The pellet was resuspended in the bath solution during patch-clamp experiments and stored at -80°C .

Expression of recombinant InsP_3R and preparation of proteoliposomes. COS-1 or HEK293 cells were transfected with the plasmid p $\text{InsP}_3\text{R-T1}$ (recombinant InsP_3R), p(1–6) $\text{InsP}_3\text{R-T1}$ ($\Delta\text{InsP}_3\text{R}$), or salmon sperm single-stranded DNA (control in Fig. 6g) as described in ref. 24. Cells were collected 48–72 h post-transfection and microsomes were prepared as described²⁴. Microsomes from control and $\text{InsP}_3\text{R1}$ -expressing cells were solubilized in 1% CHAPS buffer and fractionated through 5–20% sucrose (w/v) gradients. Gradient fractions containing InsP_3R were identified by immunoblotting with $\text{InsP}_3\text{R1}$ -specific antibody. Fresh InsP_3R fractions and parallel fractions of control cells were reconstituted into liposomes as described²⁵. Proteoliposomes from cells transfected with p $\text{InsP}_3\text{R-T1}$ yielded pronounced channel activity when incorporated into lipid bilayers, and those containing the COS-1 control had no channel activity (data not shown). HEK293 cells expressing $\Delta\text{InsP}_3\text{R1}$ were used to prepare microsomes by the procedure already described for the preparation of cerebellar microsomes.

Received 20 August; accepted 8 October 1998.

- Putney, J. W. J. & Bird, G. S. J. The inositol phosphate–calcium signaling system in nonexcitable cells. *Endocr. Rev.* **14**, 610–631 (1993).
- Berridge, M. J. Capacitative calcium entry. *Biochem. J.* **312**, 1–11 (1995).
- Parech, A. B. & Penner, R. Store depletion and calcium influx. *Physiol. Rev.* **77**, 901–930 (1997).
- Hoth, M. & Penner, R. Depletion of intracellular calcium stores activates a calcium current in mast cells. *Nature* **355**, 353–356 (1990).
- Zweifach, A. & Lewis, R. S. Mitogen-regulated Ca^{2+} current of T lymphocytes is activated by depletion of intracellular Ca^{2+} stores. *Proc. Natl Acad. Sci. USA* **90**, 6295–6299 (1993).
- Randriamampita, C. & Tsien, R. Y. Emptying of intracellular Ca^{2+} stores releases a novel small messenger that stimulates Ca^{2+} influx. *Nature* **364**, 809–814 (1993).
- Irvine, R. F. 'Quantal' Ca^{2+} release and the control of Ca^{2+} entry by inositol phosphates—a possible mechanism. *FEBS Lett.* **263**, 5–9 (1990).
- Zhu, X. *et al.* trp , a novel mammalian gene family essential for agonist-activated capacitative Ca^{2+} entry. *Cell* **58**, 661–671 (1996).
- Hurst, R. S., Zhu, X., Boulay, G., Birnbaumer, L. & Stefani, E. Ionic currents underlying Htrp3 mediated agonist-dependent Ca^{2+} influx in stably transfected HEK293 cells. *FEBS Lett.* **422**, 333–338 (1998).

- Kiselyov, K. I., Mamin, A. G., Semyonova, S. B. & Mozhayeva, G. N. Low-conductance high selective inositol (1,4,5)-trisphosphate activated Ca^{2+} channels in plasma membrane of A431 cells. *FEBS Lett.* **407**, 309–312 (1997).
- Kiselyov, K. I., Semyonova, S. B., Mamin, A. G. & Mozhayeva, G. N. Miniature Ca^{2+} channels in plasma membrane excised patches: activation by IP_3 receptor. *Pfluegers Arch.* (in the press).
- Zitt, C. *et al.* Cloning and functional expression of a human Ca^{2+} -permeable cation channel activated by calcium store depletion. *Neuron* **138**, 1333–1341 (1997).
- Zitt, C. *et al.* Expression of TRPC3 in Chinese hamster ovary cells results in calcium-activated cation currents not related to store depletion. *J. Cell Biol.* **138**, 1333–1341 (1997).
- Merritt, J. E. *et al.* SK&F 96365, a novel inhibitor of receptor-mediated calcium entry. *Biochem. J.* **271**, 515–522 (1990).
- Zhu, X., Jiang, M. & Birnbaumer, L. Receptor-activated Ca^{2+} influx via human Trp3 stably expressed in human embryonic kidney (HEK)293 cells. Evidence for a non-capacitative Ca^{2+} entry. *J. Biol. Chem.* **273**, 133–142 (1998).
- Gafni, J. *et al.* Xestospingins, potent membrane permeable blockers of the inositol 1,4,5-trisphosphate receptor. *Neuron* **19**, 723–733 (1997).
- Takemura, H., Hughes, A. R., Thastrup, O. & Putney, J. W. J. Activation of calcium entry by the tumor promoter thapsigargin in parotid acinar cells. *J. Biol. Chem.* **264**, 12266–12271 (1989).
- Smith, P. M. & Gallacher, D. V. Thapsigargin-induced Ca^{2+} mobilization in acutely isolated mouse lacrimal cells is dependent on a basal level of IP_3 and is inhibited by heparin. *Biochem. J.* **229**, 37–40 (1994).
- Sharp, A. H. *et al.* Differential immunohistochemical localization of inositol 1,4,5-trisphosphate- and ryanodine-sensitive Ca^{2+} release channels in rat brain. *J. Neurosci.* **13**, 3051–3063 (1993).
- Sugawara, H., Kurosaki, M., Takata, M. & Kurosaki, T. Genetic evidence for involvement of type 1, type 2 and type 3 inositol 1,4,5-trisphosphate receptors in signal transduction through the B-cell antigen receptor. *EMBO J.* **16**, 3078–3088 (1997).
- Beam, K. G. & Franzini-Armstrong, C. Functional and structural approaches to the study of excitation–contraction coupling. *Meth. Cell Biol.* **52**, 283–306 (1997).
- Franzini-Armstrong, C. & Jorgensen, A. O. Structure and development of E–C coupling units in skeletal muscle. *Annu. Rev. Physiol.* **56**, 509–534 (1994).
- Muallem, S. & Lee, M. G. High $[\text{Ca}^{2+}]$ domains, secretory granules and exocytosis. *Cell Calcium* **22**, 1–4 (1997).
- Sharp, A. H., Newton, C. L., Archer III, B. T. & Sudhof, T. C. Structure and expression of the rat inositol 1,4,5-trisphosphate receptor. *J. Biol. Chem.* **265**, 12579–12685 (1990).
- Ramos-Franco, J., Caenepeel, S., Fill, M. & Mignery, G. A. A single channel function of recombinant type-1 inositol 1,4,5-trisphosphate receptor ligand binding domain splice variants. *Biophys. J.* (in the press).
- Lee, M. G. *et al.* Polarized expression of Ca^{2+} channels in pancreatic and salivary cells. Correlation with initiation and propagation of $[\text{Ca}^{2+}]$ waves. *J. Biol. Chem.* **272**, 15765–15770 (1997).

Acknowledgements. We thank I. Bezprozvanny for insightful discussions. This work was supported by the NIH.

Correspondence and requests for materials should be addressed to S.M. (e-mail: smuall@mednet.swmed.edu).

The receptor Msn5 exports the phosphorylated transcription factor Pho4 out of the nucleus

Arie Kaffman, Nicole Miller Rank, Elizabeth M. O'Neill, Linda S. Huang & Erin K. O'Shea

Department of Biochemistry and Biophysics, University of California at San Francisco, School of Medicine, San Francisco, California 94143-0448, USA

The movement of many transcription factors, kinases and replication factors between the nucleus and cytoplasm is important in regulating their activity¹. In some cases, phosphorylation of a protein regulates its entry into the nucleus²; in others, it causes the protein to be exported to the cytoplasm^{3–6}. The mechanism by which phosphorylation promotes protein export from the nucleus is poorly understood. Here we investigate how the export of the yeast transcription factor Pho4 is regulated in response to changes in phosphate availability. We show that phosphorylation of Pho4 by a nuclear complex of a cyclin with a cyclin-dependent kinase, Pho80–Pho85, triggers its export from the nucleus. We also find that the shuttling receptor used by Pho4 for nuclear export is the importin- β -family member Msn5 (refs 7, 8), which is required for nuclear export of Pho4 *in vivo* and binds only to phosphorylated Pho4 in the presence of the GTP-bound form of yeast Ran *in vitro*. Our results reveal a simple mechanism by which phosphorylation can control the nuclear export of a protein.

When yeast are grown in phosphate-rich medium, the transcription factor Pho4 is phosphorylated by the Pho80–Pho85 cyclin–cyclin-dependent kinase (CDK) complex⁹ and is localized to the cytoplasm¹⁰, thereby turning off transcription of genes that are turned on specifi-

cally by phosphate starvation¹¹. When yeast are starved of phosphate, the CDK inhibitor Pho81 inactivates Pho80–Pho85 (refs 12, 13), and Pho4 is not phosphorylated. Unphosphorylated Pho4, but not phosphorylated Pho4, is recognized by the non-classical import receptor Pse1/Kap121, indicating that nuclear import of Pho4 may be regulated in response to phosphate availability¹⁴.

For several reasons the regulation of Pho4 import alone is unlikely to be sufficient to account for the change in Pho4 localization that is induced by phosphorylation. First, the Pho80–Pho85 cyclin–CDK complex localizes to the nucleus (Fig. 1a), whereas phosphorylated Pho4 is cytoplasmic¹⁰. These observations indicate that Pho4 may be phosphorylated in the nucleus and exported to the cytoplasm.

Second, we observed that phosphorylated Pho4 was rapidly exported from the nucleus. To monitor export of Pho4, we first grew a yeast strain expressing a fusion of Pho4 and green fluorescent protein (GFP) in medium lacking phosphate; this led to accumulation of unphosphorylated Pho4–GFP in the nucleus (Fig. 1b, top). We added cycloheximide to prevent further protein synthesis. We then added phosphate to activate the Pho80–Pho85 kinase, and monitored the localization of Pho4–GFP. After 3–6 min the localization of Pho4–GFP was indistinguishable from that seen in cells grown in phosphate-rich medium (Fig. 1b, top). If no phosphate was added to the culture, Pho4–GFP remained in the nucleus for several hours (data not shown). An alternative explanation for these results, that phosphorylation of Pho4–GFP leads to degradation of nuclear Pho4–GFP followed by cytoplasmic accumulation of newly synthesized protein, is unlikely because protein synthesis was inhibited. To test the role of phosphorylation in the rapid export of Pho4, we studied the export of Pho4^{SA}, a mutant Pho4 containing five serine-to-alanine substitutions at the sites of phosphorylation by Pho80–Pho85 (ref. 10). Pho4^{SA} did not leave the nucleus upon

addition of phosphate to a starved culture of yeast (Fig. 1b, bottom). In addition, in strains lacking either the cyclin Pho80 or the CDK Pho85, Pho4–GFP did not leave the nucleus upon addition of phosphate (data not shown). These data indicate that phosphorylation of Pho4 is required for its cytoplasmic accumulation.

To determine whether the export of Pho4 from the nucleus is regulated by phosphorylation, it is necessary to examine export in the absence of import¹⁵. If the export of Pho4 is enhanced by phosphorylation, phosphorylated Pho4 will accumulate in the cytoplasm more rapidly than unphosphorylated Pho4 when import is blocked. In contrast, if the export of Pho4 is not regulated by phosphorylation, both forms of Pho4 will accumulate in the cytoplasm at the same rate when import is blocked (Fig. 2a).

We first demonstrated that two temperature-sensitive yeast mutants with a conditional defect in protein import, *nsp1^{ts}* (refs 16, 17) and *nup49-313* (refs 16, 18), were defective for import of Pho4 when grown at the restrictive temperature (data not shown). An experiment using the *nsp1^{ts}* strain expressing Pho4–GFP was performed as follows (Fig. 2b, top). First, the yeast were starved for phosphate at 23 °C, allowing unphosphorylated Pho4–GFP to accumulate in the nucleus. The culture was then shifted to 38.5 °C to impair protein import, treated with cycloheximide and split in half. Phosphate was added to half of the culture to activate Pho80–Pho85 and trigger phosphorylation of Pho4, and was not added to the other half. Pho4–GFP was rapidly transported to the cytoplasm in the culture to which phosphate had been added, whereas it remained nuclear in the untreated culture for several more hours (Fig. 2b, bottom). In addition, when phosphate was added, Pho4^{SA} remained in the nucleus for several hours when import was blocked (data not shown). These data indicate that phosphorylation may regulate nuclear export of Pho4.

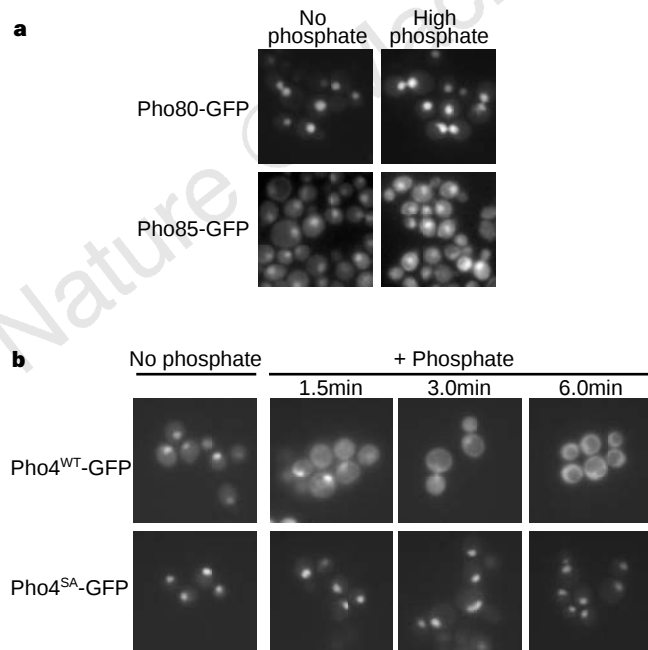


Figure 1 Phosphorylation of Pho4 by nuclear Pho80–Pho85 promotes its rapid export from the nucleus. **a**, Direct fluorescence microscopy of cells carrying a plasmid expressing Pho80–GFP or Pho85–GFP grown in high- or no-phosphate media. As both Pho80 and Pho85 are required to phosphorylate Pho4 (ref. 9), the localization of Pho80 restricts the kinase activity to the nucleus. **b**, Fluorescence microscopy of cells carrying a plasmid expressing Pho4^{WT}-GFP or Pho4^{SA}-GFP (a non-phosphorylatable mutant) under the control of the *PHO4* promoter. Cells were starved of phosphate (no phosphate) and incubated with cycloheximide for 10 min. At *t* = 0 min phosphate was added to the culture and localization of Pho4–GFP was monitored as a function of time.

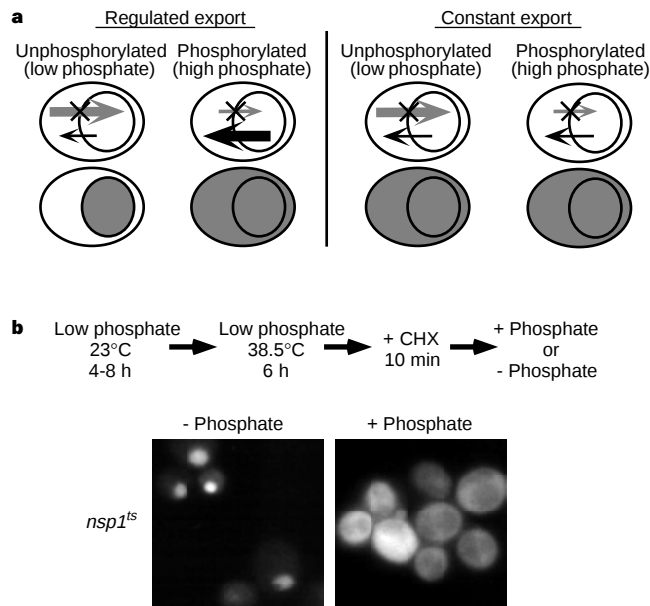


Figure 2 The rate of export of Pho4 from the nucleus is regulated by phosphorylation. **a**, Predicted outcomes of blocking import when export is either regulated or is constant. **b**, Top, outline of the experiment used to determine whether the rate of export is regulated in response to phosphorylation; CHX, cycloheximide. Bottom, localization of Pho4^{WT}-GFP in an *nsp1^{ts}* strain^{16,17} that has been starved of phosphate and incubated in cycloheximide at 38.5 °C. The culture was split and phosphate was added to half of the culture (+ Phosphate) but not to the other half (– Phosphate). The addition of phosphate to the *nsp1^{ts}* strain resulted in export of Pho4–GFP with the same kinetics as seen in the wild-type strain.

To understand the molecular mechanism by which phosphorylation regulates Pho4 export, we wished to identify the Pho4 export receptor. We therefore studied the localization of Pho4-GFP in yeast strains containing mutations in putative receptors involved in nuclear transport. Thirteen proteins in yeast have homology to importin- β , a receptor that recognizes nuclear-localization signals (NLSs) on proteins imported into the nucleus^{7,8}. Some importin- β -family members are transport receptors; others have not yet been characterized. No defect in Pho4 export was observed in yeast strains containing mutations in the importin- β -family members Kap104 (ref. 19), Cse1 (ref. 20), Xpo1 (ref. 21), Sxm1 (ref. 22), Kap123 (refs 22, 23) and Mtr10 (ref. 24). However, in a strain lacking the importin- β -family member Msn5/Ste21, Pho4-GFP could not be exported and was localized constitutively to the nucleus (Fig. 3a). The defect in Pho4 localization in the *msn5* Δ strain is not an indirect result of a defect in the phosphate signal-transduction pathway, as Pho4 is still phosphorylated normally in this strain (Fig. 3b). Although Pho4-GFP was localized to the nucleus in the *msn5* Δ strain, expression of the secreted acid phosphatase Pho5, which is activated by Pho4 under conditions of phosphate starvation, was not induced because there are extra mechanisms for regulating the transcriptional activity of Pho4 (E.K.O. *et al.*, manuscript in preparation).

One model that may explain regulated export of Pho4 is that Msn5 preferentially binds the phosphorylated form of Pho4. We tested this hypothesis by performing an *in vitro* binding experiment. As all protein-export pathways characterized so far require the small GTPase Ran in the GTP-bound state^{25–28}, we tested whether Ran-GTP is required for Msn5 to bind Pho4. Phosphorylated Pho4 bound to Msn5 in yeast extract when a Ran mutant was added that is locked in the GTP-bound state (human RanQ69L-GTP)²⁹ (Fig. 4a, top, lane 1). In the absence of RanQ69L-GTP, no interaction was

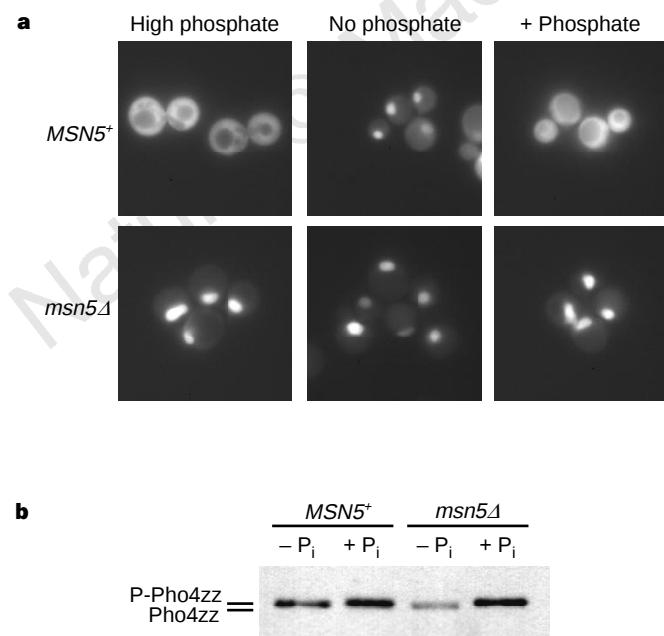


Figure 3 Msn5 is required for export of Pho4. **a**, Localization of Pho4^{WT}-GFP in wild-type (*MSN5⁺*) and *msn5* Δ strains grown in high-phosphate medium or no-phosphate medium or 10 min after the addition of phosphate to phosphate-starved cells (+ Phosphate). **b**, Immunoblot detecting phosphorylation of Pho4^{WT}-zz in extracts from *MSN5⁺* and *msn5* Δ strains. Yeast strains were grown in low-phosphate medium, and phosphate was either added (+ P_i) or not added (- P_i) to the culture. Addition of phosphate to *MSN5⁺* and *msn5* Δ strains expressing Pho4^{WT}-zz grown in phosphate-depleted medium resulted in a shift in the electrophoretic mobility of Pho4, indicating that Pho4 was phosphorylated (P-Pho4zz)⁹.

observed between phosphorylated Pho4 and Msn5 (Fig. 4a, top, lane 2). Pho4 must be phosphorylated to interact with Msn5, as, even in the presence of RanQ69L-GTP, no interaction was seen between Msn5 and the mutant Pho4^{SA}, which cannot be phosphorylated (Fig. 4a, top, lane 3).

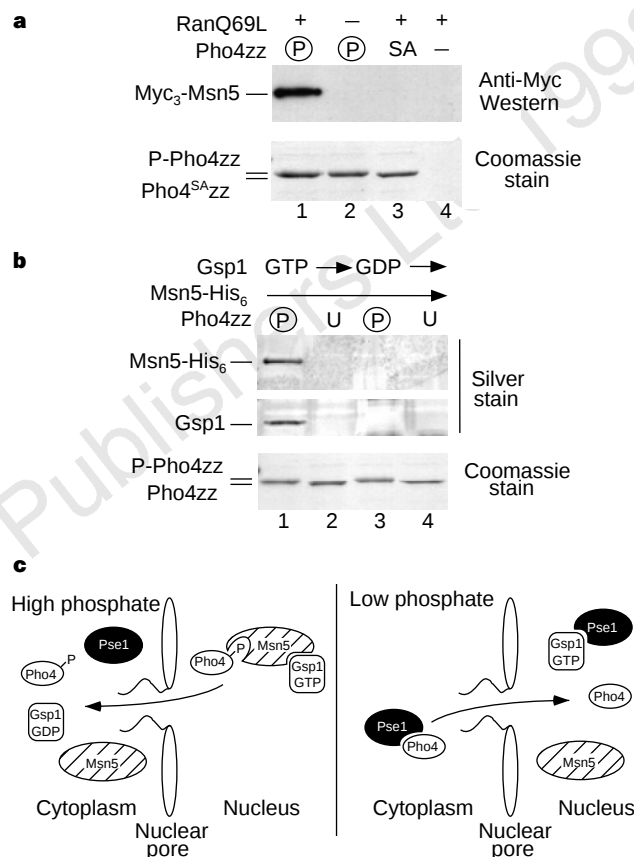


Figure 4 Msn5 binds directly to phosphorylated Pho4 in the presence of Ran-GTP. **a**, *In vitro*-phosphorylated Pho4^{WT}-zz (circ P) or Pho4^{SA}-zz (SA) or a no Pho4 control (-) was purified using IgG-Sepharose beads, and incubated with extract containing Myc₃-tagged Msn5 in the presence or absence of RanQ69L (ref. 29). Top, bound proteins were eluted with 1 M MgCl₂, concentrated using methanol-chloroform precipitation, separated on 7.8% SDS-PAGE and immunoblotted with anti-Myc antibodies. Lane 1, phosphorylated Pho4^{WT}-zz (P-Pho4zz) with RanQ69L; lane 2, phosphorylated Pho4^{WT}-zz without RanQ69L; lane 3, phosphorylated Pho4^{SA}-zz with RanQ69L; lane 4, beads alone with RanQ69L (lane 4). Bottom, to ensure that the same amounts of Pho4-zz were immobilized on IgG-Sepharose in the lanes shown in the top panel, Pho4-zz was eluted with 0.5 M acetic acid, pH 3.4, separated on 7.8% SDS-PAGE, and visualized with Coomassie blue. Lanes are as described above. **b**, Pho4^{WT}-zz was phosphorylated (circled P) or mock-phosphorylated (U) *in vitro*, and purified using IgG-Sepharose beads. Immobilized proteins were incubated with 1 μ M purified Msn5-His₆ in the presence of 1 μ M purified Myc-Gsp1 loaded with either GTP or GDP. Top, bound proteins were eluted with 1 M MgCl₂, concentrated using methanol-chloroform precipitation, separated on 7.8% SDS-PAGE and visualized with silver staining. Lane 1, phosphorylated Pho4^{WT}-zz with Myc-Gsp1 loaded with GTP; lane 2, unphosphorylated Pho4^{WT}-zz with Myc-Gsp1 loaded with GTP; lane 3, phosphorylated Pho4^{WT}-zz with Myc-Gsp1 loaded with GDP; lane 4, unphosphorylated Pho4^{WT}-zz with Myc-Gsp1 loaded with GDP. Roughly 20% of the Msn5-His₆ bound to phosphorylated Pho4 in the presence of Gsp1-GTP. Bottom, to ensure that the amount of Pho4-zz immobilized on IgG-Sepharose did not vary between lanes, Pho4^{WT}-zz was eluted with 0.5 M acetic acid, pH 3.4, separated on 7.8% SDS-PAGE, and visualized with Coomassie blue. Lanes are as described above. **c**, Model for the regulated nucleocytoplasmic localization of Pho4 in response to extracellular phosphate.

There are several models that could explain the preferential binding of Msn5 to phosphorylated Pho4. Preferential binding might require an adapter protein to enable the interaction with phosphorylated Pho4, or a protein that masks the Msn5-binding site on unphosphorylated Pho4, or it might be a direct consequence of the ability of Msn5 to distinguish between the two forms of Pho4. To determine whether Msn5 alone can bind preferentially to the phosphorylated form of Pho4, we used purified recombinant proteins to test the binding of Pho4 to Msn5 and the yeast Ran homologue Gsp1. Purified Msn5 bound preferentially to phosphorylated Pho4 in the presence of Gsp1 loaded with GTP (Fig. 4b, top, lane 1), but not when Gsp1 was loaded with GDP (Fig. 4b, top, lane 3). Purified Msn5 did not interact with unphosphorylated Pho4, even in the presence of Gsp1-GTP (Fig. 4b, top, lane 2). These data indicate that no additional proteins are required for Msn5 to bind preferentially to phosphorylated Pho4 and Gsp1, and that the interaction between Pho4 and Msn5 requires Gsp1-GTP. The observation that Gsp1-GTP bound to Pho4 in the presence of Msn5 (Fig. 4b, middle) but not in its absence (data not shown) indicates that phosphorylated Pho4, Gsp1-GTP and Msn5 may form a ternary complex. These results are consistent with the properties of other characterized complexes of export receptor and cargo, which require the export receptor, cargo and Ran-GTP for stable complex formation^{25–28}.

Our data indicate a model that may explain how localization of Pho4 is regulated by phosphorylation (Fig. 4c). In high-phosphate medium, the activated Pho80-Pho85 cyclin-CDK complex phosphorylates Pho4 in the nucleus, triggering association between phosphorylated Pho4, the export receptor Msn5 and Gsp1-GTP. This trimeric complex is then rapidly exported from the nucleus. By analogy to what has been shown for the other export receptors^{25–28}, we propose that the Pho4-Msn5-Gsp1-GTP complex is disassembled in the cytoplasm by hydrolysis of GTP by Gsp1, stimulated by Rna1 (yeast RanGAP, a GTPase-activating protein) and Yrb1 (yeast RanBP1, a Ran-binding protein). Phosphorylated cytoplasmic Pho4 is then prevented from re-entering the nucleus because phosphorylation of Pho4 inhibits interaction with the import receptor Pse1 (ref. 14). We have determined that the phosphorylation sites necessary and sufficient for regulating export are distinct from the site that regulates Pho4 import (E.K.O. *et al.*, manuscript in preparation). The dual role for phosphorylation in promoting export and preventing re-import is likely to be a general mechanism, as it allows cells to rapidly and efficiently regulate gene expression in response to environmental signals.

The ability of Msn5 to preferentially recognize phosphorylated Pho4 may be explained in two ways. The Msn5-binding site present on phosphorylated Pho4 could be revealed by a phosphorylation-induced conformational change. Alternatively, Msn5 may recognize a phosphopeptide within Pho4. In this case, Msn5 may also function as an export receptor for other phosphorylated proteins. The existence of an export receptor dedicated to phosphorylated cargo would provide a simple mechanism for selectively exporting proteins that are phosphorylated in response to cellular signals.

Although many proteins are regulated by phosphorylation-induced export from the nucleus, the molecular mechanism by which phosphorylation induces nuclear export has not been understood previously. Our results reveal a new pathway for nuclear export that uses an export receptor that requires only Ran-GTP to bind exclusively to the phosphorylated form of its cargo. These results provide a paradigm for understanding how phosphorylated proteins are exported from the nucleus. □

Methods

Plasmids and strains. *PHO4*, *PHO80* and *PHO85* were deleted in strain K699 (*MATa ade2-1 trp1-1 can1-100 leu2-3,112 his3-11,15 ura3*) and *MSN5* and *PHO4* were deleted in strain IH3195 (*MATa HMLa HMRa ura3-52 ade2-101 met1 leu2Δ1 trp1Δ99*), which is isogenic to JC2-1B (ref. 30), by standard gene replacement. *nup49-313* and the isogenic *NUP49⁺* strain^{16,18}, and *nsp1^{ts}* and its

isogenic strain^{16,17} have been described previously. High-phosphate cultures were grown in standard synthetic dropout (SD) medium; low-phosphate cultures were grown in phosphate-depleted SD medium; and no-phosphate cultures were grown in SD medium with KCl replacing KH_2PO_4 (ref. 14). The construction of plasmids expressing Pho4^{WT} and Pho4^{SA} fused to either GFP or to the zz tag from Protein A, and of the plasmid expressing Myc-Gsp1, has been described¹⁴. Pho80-GFP and Pho85-GFP were expressed from their own promoter on a high-copy or low-copy plasmid, respectively. Both plasmids complement the phenotype of the respective deletion strain. Detailed descriptions of plasmids used in this study are available on request.

Detection of phosphorylation of Pho4-zz. Isogenic *pho4Δ* and *pho4Δmsn5Δ* strains expressing Pho4^{WT}-zz from a plasmid were inoculated at an A_{600} value of 0.1–0.3 and grown in low-phosphate medium for 8 h. Cells were collected 10 min after the addition of KH_2PO_4 or KCl to a final concentration of 20 mM and were frozen in liquid nitrogen. Samples were lysed by boiling for 5 min in the presence of SDS-PAGE loading buffer followed by bead-beating with acid-washed 0.5-mm glass beads for 5 min. Samples were reboiled for 3 min and 10–15 μg total protein was separated using 7.8% SDS-PAGE and immunoblotted to detect the zz tag.

Phosphate starvation and microscopy. Phosphate starvation and microscopy were done as described¹⁴, except that the photographs for Fig. 2 were taken using an Olympus PM-30 camera.

Examination of Pho4^{WT}-GFP in the *nsp1^{ts}* strain. The *nsp1^{ts}* strain carrying a plasmid expressing Pho4^{WT}-GFP was starved for phosphate for 4–8 h in low-phosphate medium at 23 °C, resulting in the accumulation of Pho4^{WT}-GFP in the nucleus. Cells were then shifted to 38.5 °C for 4–6 h to impair protein import. Cycloheximide was added to a final concentration of 0.1 mg ml⁻¹ and the culture was incubated for 10 min. For the phosphate-containing samples, 4.5 μl culture plus 0.5 μl of 15 mg ml⁻¹ KH_2PO_4 were mixed on a slide and Pho4-GFP localization was examined by direct fluorescence. Similar results were obtained with the *nup49-313* strain (data not shown).

Expression and purification of recombinant proteins. Msn5 was expressed in *Escherichia coli* with a His₆ tag at the carboxy terminus essentially as described for Pse1-His₆ (ref. 14), with the following modifications: cells were lysed in lysis buffer containing 10 mM imidazole and Msn5-His₆ was eluted with a 10–1,000 mM imidazole gradient. Myc-Gsp1, Pho4^{WT}-zz and Pho4^{SA}-zz were expressed and purified as described¹⁴. RanQ69L was expressed in strain BL21 grown at 37 °C in Luria broth + 100 μg ml⁻¹ carbenicillin to an A_{600} value of 0.4 and induced with 0.4 mM isopropyl-β-D-thiogalactoside at 30 °C for 1.5 h. RanQ69L was purified as reported for Myc-Gsp1 (ref. 14).

Binding assay. *In vitro* phosphorylation of Pho4-zz and loading of Myc-Gsp1 with GTP or GDP were done as described¹⁴. Yeast whole cell extract was prepared nearly as described¹⁴ except that phosphatase inhibitors were included (80 mM β-glycerophosphate, 10 mM NaF, 10 mM calyculin A). Binding of Myc₃-Msn5 to Pho4-zz was essentially as described for Pse1 (ref. 14), with the following modifications: first, purified RanQ69L was added to a final concentration of 5 μM; second, an ATP-regenerating system was added (1 mM ATP, 1 mM GTP, 50 μg ml⁻¹ creatine kinase, 5 mM creatine phosphate); and third, phosphatase inhibitors were reduced to 40 mM β-glycerophosphate, 5 mM NaF and 5 mM calyculin A. Binding with purified components was done by incubating 25 μg purified Pho4-zz immobilized on IgG-Sepharose beads with a final concentration of 1 μM Myc-Gsp1 and 1 μM Msn5-His₆ in the presence of 0.5 mg ml⁻¹ bovine serum albumin for 1.5 h at 4 °C.

Received 15 September; accepted 15 October 1998.

1. Nigg, E. A. Nucleocytoplasmic transport: signals, mechanisms and regulation. *Nature* **386**, 779–787 (1997).
2. Jans, D. A. & Hubner, S. Regulation of protein transport to the nucleus: central role of phosphorylation. *Physiol. Rev.* **76**, 651–685 (1996).
3. Timmerman, L. A., Clipstone, N. A., Ho, S. N., Northrop, J. P. & Crabtree, G. R. Rapid shuttling of NF-AT in discrimination of Ca^{2+} signals and immunosuppression. *Nature* **383**, 837–840 (1996).
4. Shibasaki, F., Price, E. R., Milan, D. & McKeon, F. Role of kinases and the phosphatase calcineurin in the nuclear shuttling of transcription factor NF-AT4. *Nature* **382**, 370–373 (1996).
5. De Vit, M. J., Waddle, J. A. & Johnston, M. Regulated nuclear translocation of the Mig1 glucose repressor. *Mol. Biol. Cell* **8**, 1603–1618 (1997).
6. Gorner, W. *et al.* Nuclear localization of the C2H2 zinc finger protein Msn2p is regulated by stress and protein kinase A activity. *Genes Dev.* **12**, 586–597 (1998).
7. Fornerod, M. *et al.* The human homologue of yeast CRM1 is in a dynamic subcomplex with CAN/Nup214 and a novel nuclear pore component Nup88. *EMBO J.* **16**, 807–816 (1997).
8. Gorlich, D. *et al.* A novel class of RanGTP binding proteins. *J. Cell Biol.* **138**, 65–80 (1997).
9. Kaffman, A., Herskowitz, I., Tjian, R. & O'Shea, E. K. Phosphorylation of the transcription factor PHO4 by a cyclin-CDK complex, PHO80-PHO85. *Science* **263**, 1153–1156 (1994).
10. O'Neill, E. M., Kaffman, A., Jolly, E. R. & O'Shea, E. K. Regulation of PHO4 nuclear localization by the PHO80-PHO85 cyclin-CDK complex. *Science* **271**, 209–212 (1996).

11. Oshima, Y. The phosphatase system in *Saccharomyces cerevisiae*. *Genes Genet. Syst.* **72**, 323–334 (1997).
12. Ogawa, N. *et al.* Functional domains of Pho81p, an inhibitor of Pho85p protein kinase, in the transduction pathway of Pi signals in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **15**, 997–1004 (1995).
13. Schneider, K. R., Smith, R. L. & O'Shea, E. K. Phosphate-regulated inactivation of the kinase PHO80–PHO85 by the CDK inhibitor PHO81. *Science* **266**, 122–126 (1994).
14. Kaffman, A., Rank, N. M. & O'Shea, E. K. Phosphorylation regulates association of the transcription factor Pho4 with its import receptor Pse1/Kap121. *Genes Dev.* **12**, 2673–2683 (1998).
15. Lee, M. S., Henry, M. & Silver, P. A. A protein that shuttles between the nucleus and the cytoplasm is an important mediator of RNA export. *Genes Dev.* **10**, 1233–1246 (1996).
16. Wimmer, C., Doye, V., Grandi, P., Nehrass, U. & Hurt, E. C. A new subclass of nucleoporins that functionally interact with nuclear pore protein NSP1. *EMBO J.* **11**, 5051–5061 (1992).
17. Nehrass, U. *et al.* NSP1: a yeast nuclear envelope protein localized at the nuclear pores exerts its essential function by its carboxy-terminal domain. *Cell* **61**, 979–989 (1990).
18. Doye, V., Wepf, R. & Hurt, E. C. A novel nuclear pore protein Nup133p with distinct roles in poly(A)⁺ RNA transport and nuclear pore distribution. *EMBO J.* **13**, 6062–6075 (1994).
19. Aitchison, J. D., Blobel, G. & Rout, M. P. Kap104p: a karyopherin involved in the nuclear transport of messenger RNA binding proteins. *Science* **274**, 624–627 (1996).
20. Xiao, Z., McGrew, J. T., Schroeder, A. J. & Fitzgerald-Hayes, M. CSE1 and CSE2, two new genes required for accurate mitotic chromosome segregation in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **13**, 4691–4702 (1993).
21. Stade, K., Ford, C. S., Guthrie, C. & Weiss, K. Exportin 1 (Crm1p) is an essential nuclear export factor. *Cell* **90**, 1041–1050 (1997).
22. Seedorf, M. & Silver, P. A. Importin/karyopherin protein family members required for mRNA export from the nucleus. *Proc. Natl Acad. Sci. USA* **94**, 8590–8595 (1997).
23. Rout, M. P., Blobel, G. & Aitchison, J. D. A distinct nuclear import pathway used by ribosomal proteins. *Cell* **89**, 715–725 (1997).
24. Kadowaki, T. *et al.* Isolation and characterization of *Saccharomyces cerevisiae* mRNA transport-defective (mtr) mutants. *J. Cell Biol.* **126**, 649–659 (1994).
25. Arts, G. J., Fornerod, M. & Mattaj, J. W. Identification of a nuclear export receptor for tRNA. *Curr. Biol.* **8**, 305–314 (1998).
26. Kutay, U., Bischoff, F. R., Kostka, S., Kraft, R. & Gorlich, D. Export of importin alpha from the nucleus is mediated by a specific nuclear transport factor. *Cell* **90**, 1061–1071 (1997).
27. Kutay, U. *et al.* Identification of a tRNA-specific nuclear export receptor. *Mol. Cell* **1**, 359–369 (1998).
28. Fornerod, M., Ohno, M., Yoshida, M. & Mattaj, J. W. CRM1 is an export receptor for leucine-rich nuclear export signals. *Cell* **90**, 1051–1060 (1997).
29. Bischoff, F. R., Klebe, C., Kretschmer, J., Wittinghofer, A. & Ponstingl, H. RanGAP1 induces GTPase activity of nuclear Ras-related Ran. *Proc. Natl Acad. Sci. USA* **91**, 2587–2591 (1994).
30. Chenevert, J., Valtz, N. & Herskowitz, I. Identification of genes required for normal pheromone-induced cell polarization in *Saccharomyces cerevisiae*. *Genetics* **136**, 1287–1296 (1994).

Acknowledgements. We thank I. Herskowitz, J. Li, J. Weissman and members of the O'Shea laboratory for comments on the manuscript; R. Bischoff for the RanQ69L plasmid; M. Lenburg for the Pho85–GFP plasmid; and J. Aitchison, K. Weiss, P. Silver, M. Fitzgerald-Hayes, E. Hurt and A. Tartakoff for yeast strains. A.K. is a Fellow of the UCSF Medical Scientist Training Program. N.M.R. and L.S.H. were supported by fellowships from the NIH and E.M.O. was supported by a fellowship from the Jane Coffin Childs Foundation. This work was supported by the David and Lucile Packard Foundation and by an NSF Presidential Faculty Fellowship (E.K.O.).

Correspondence and requests for materials should be directed to E.K.O. (e-mail: oshea@biochem.ucsf.edu).

Crystal structure of the ligand-binding domain of the receptor tyrosine kinase EphB2

Juha-Pekka Himanen*, Mark Henkemeyer† & Dimitar B. Nikolov*

* Cellular Biochemistry and Biophysics Program, Memorial-Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10021, USA

† Center for Developmental Biology, University of Texas Southwestern Medical Center, Dallas, Texas 75235-9133, USA

The Eph receptors, which bind a group of cell-membrane-anchored ligands known as ephrins, represent the largest sub-family of receptor tyrosine kinases (RTKs)¹. They are predominantly expressed in the developing and adult nervous system² and are important in contact-mediated axon guidance^{3–6}, axon fasciculation^{5,7} and cell migration^{8–11}. Eph receptors are unique among other RTKs in that they fall into two subclasses with distinct ligand specificities¹², and in that they can themselves function as ligands to activate bidirectional cell–cell signalling^{4,13,14}. We report here the crystal structure at 2.9 Å resolution of the amino-terminal ligand-binding domain of the EphB2 receptor (also known as Nuk)^{15–17}. The domain folds into a compact jellyroll β-sandwich composed of 11 antiparallel β-strands. Using structure-based mutagenesis, we have identified an extended loop that is important for ligand binding and class

specificity. This loop, which is conserved within but not between Eph RTK subclasses, packs against the concave β-sandwich surface near positions at which missense mutations cause signalling defects¹⁸, localizing the ligand-binding region on the surface of the receptor.

The extracellular region of Eph receptors consists of two fibronectin type III repeats, a cysteine-rich region, and a conserved 180-amino-acid N-terminal 'globular' domain (Fig. 1) which is both necessary and sufficient for bindings of the receptors to their ephrin ligands¹⁶. Eph receptors bind their ephrin ligands with high affinity ($K_d = 0.5–15.0$ nM) and with one-to-one stoichiometry^{16,17,19} (Fig. 2).

We determined the structure of the biologically active (Fig. 2) ligand-binding domain of the murine EphB2 receptor by X-ray crystallography using the multiple isomorphous replacement (MIR) method (Table 1 and Figs 3, 4). Our model is refined at 2.9 Å resolution to an *R*-factor of 20.6% with tightly restrained temperature factors and good stereochemistry. The domain has dimensions of roughly 50 Å × 40 Å × 30 Å. It has two antiparallel β-sheets, with the usual left-handed twist, packed against each other to form a compact β-sandwich, and a short 3₁₀ helix. The concave β-sheet is composed of strands C, F, F', L, H and I, and the convex β-sheet of strands D, E, A, M, G, K and J (Fig. 4c). The extensive hydrophobic core created by approximation of the two β-sheets and the short 3₁₀ helix contains no cavities and is dominated by aromatic side chains. The β-strands are connected by loops of varying length, including a long loop (between strands H and I; coloured in orange and red in Fig. 4a) that packs against the full width of the concave β-sheet, and two loops (D–E and J–K) that protrude from the two sides of the β-sandwich, closing on the middle of the convex β-sheet. The H–I loop (in the front in Fig. 4a) is well-ordered, as are the loops running across the top, whereas the D–E and J–K loops at the back are disordered, with several residues that cannot be located in the electron-density map. Two disulphide bridges stabilize the loops at the top of the β-sandwich: Cys 105 binds Cys 115 in the longest of the top loops (G–H), and Cys 70, the first residue of loop E–F, binds Cys 192, the last residue of loop L–M.

The EphB2 ligand-binding domain has a jellyroll folding topology²⁰ (Fig. 4c). Comparison of this structure with the contents of the FSSP database²¹ reveals considerable similarity with the carbohydrate-binding domains of glucanases, legume lectins, β-galactosidases, sialidases, cellulases, bacterial toxins and influenza virus haemagglutinin. This raises two interesting possibilities, namely that the homology in molecular architecture includes the location of the ligand-binding site, or that the carbohydrate moieties at the putative glycosylation sites of the ephrins³ are directly involved in ligand–receptor recognition. Our results support the first possibility (see below); however, we suggest that the Eph receptors are not lectins. Indeed, the purified EphB2 globular domain forms a stable, high-affinity complex with a bacterially expressed non-glycosylated ephrin-B2 ligand preparation (Fig. 2) (K_d is within the range of values reported for recombinant glycosylated ephrins^{2,16,17,19}). Furthermore, addition of a tenfold molar excess of *N*-linked type oligosaccharides (Oxford GlycoSystem) has no effect on the *in vitro* Eph-receptor–ephrin interaction (data not shown). Beyond the similarity in overall topology between the EphB2 ligand-binding domain and the carbohydrate-binding proteins, the precise structures are quite different (Fig. 4d) and the best structural alignments include stretches of 100–120 amino acids with root mean square deviations between α-carbon positions of 3.0 Å.

The ligand-binding domain of Eph receptors is unique to this family of RTKs²² and shares no significant amino-acid-sequence homology with other known proteins. Nevertheless, jellyroll folding topology is observed in the extracellular domains of other proteins including, in addition to the carbohydrate-binding proteins, the tumour-necrosis factor (TNF) family²³ (which includes TNF, lym-

Table 1 Crystallographic analysis

Crystal	Native	Sm1	Sm2	Os1	Os2	Au	Pt1	Pt2
Resolution (Å)	2.9	3.2	3.0	3.0	3.2	3.0	3.2	3.0
Redundancy	5.0	12.8	5.3	4.8	3.6	4.7	3.7	4.5
Data coverage (%)	95.1	94.0	98.2	90.7	90.1	89.4	80.6	89.3
R_{sym} (%)	11.4	17.1	10.9	12.2	14.4	7.6	10.1	6.5
Multiple isomorphous replacement analysis (15.0–3.2 Å)								
Mean fractional isomorphous difference (%)†		19.5	19.4	19.2	20.7	13.7	15.1	15.2
Binding sites		3	3	3	3	4	3	7
Phasing power‡		2.07	2.13	1.60	1.31	0.84	0.96	1.00
R_{CS}		0.64	0.64	0.69	0.77	0.83	0.84	0.84
Mean overall figure of merit	0.72							
Refinement	Resolution (Å)	Reflections ($ F > 2.0\sigma(F)$) working/test	Number of atoms	R value $R_{\text{cryst}}/R_{\text{free}}$	Bonds (Å)	R.m.s. deviations Angles (°)	B -factors (Å ²)	
	8.0–2.9	4,214/340	1,377	20.6/31.4	0.012	1.97	1.93	

The space group is $P4_22_1$ ($a = b = 55$ Å, $c = 159$ Å) with one molecule in the asymmetric unit. The heavy-metal compounds are Sm1 = Sm2 = samarium (III) acetate trihydrate, Os1 = Os2 = potassium hexachloroosmate (IV), Au = sodium tetrachloroaurate (III), Pt1 = diamine dichloroplatinate (II), and Pt2 = potassium tetrachloroplatinate (II).

* $R_{\text{sym}} = \sum |I - \langle I \rangle| / \sum I$, where I is the observed intensity and $\langle I \rangle$ is the average intensity obtained from multiple observations of symmetry-related reflections.

† Mean fractional isomorphous difference = $\sum ||F_{\text{PH}}| - |F_{\text{P}}|| / \sum |F_{\text{P}}|$ where $|F_{\text{P}}|$ is the protein structure-factor amplitude and $|F_{\text{PH}}|$ is the heavy-atom-derivative structure-factor amplitude.

‡ Phasing power = root-mean-square ($|F_{\text{H}}|/E$), where $|F_{\text{H}}|$ is the heavy-atom structure-factor amplitude and E is the residual lack of closure.

§ $R_{\text{CS}} = \sum ||F_{\text{H(obs)}}| - |F_{\text{H(calc)}}|| / \sum |F_{\text{H(obs)}}|$, where $|F_{\text{H(obs)}}|$ is the observed heavy-atom structure-factor amplitude and $|F_{\text{H(calc)}}|$ is the calculated heavy-atom structure-factor amplitude.

||r.m.s. deviations in bond lengths and angles are the respective root-mean-square deviations from ideal values. The r.m.s. thermal parameter (B -factor) is the r.m.s. deviation between the B -values of covalently bound atomic pairs.

phototoxin, CD40 ligand and complement C1q) and several viral coat proteins. The TNF-family members are all trimeric and bind their receptors on the 'side' of the jellyroll, at the interface between trimer subunits²³. Eph receptors, on the other hand, are monomeric (including the full-length recombinant extracellular region)^{16,19}, and therefore their functional ligand-binding site is different from that of TNF proteins.

Discussion of the precise mechanisms of ligand recognition by the Eph RTKs must await high-resolution structures of appropriate

ephrin–Eph-receptor complexes. However, the structure of a functional ligand-binding core domain does allow us to predict the location of the ligand-recognition site on the receptor surface and to test our hypothesis using structure-based mutagenesis. Several results (see below) indicate that the ligand-binding surface of EphB2 is localized to the upper part of the concave β -sheet, including the H–I loop traversing the middle region of the structure as well as the loops running across the upper surface of the β -sandwich (Fig. 4d).

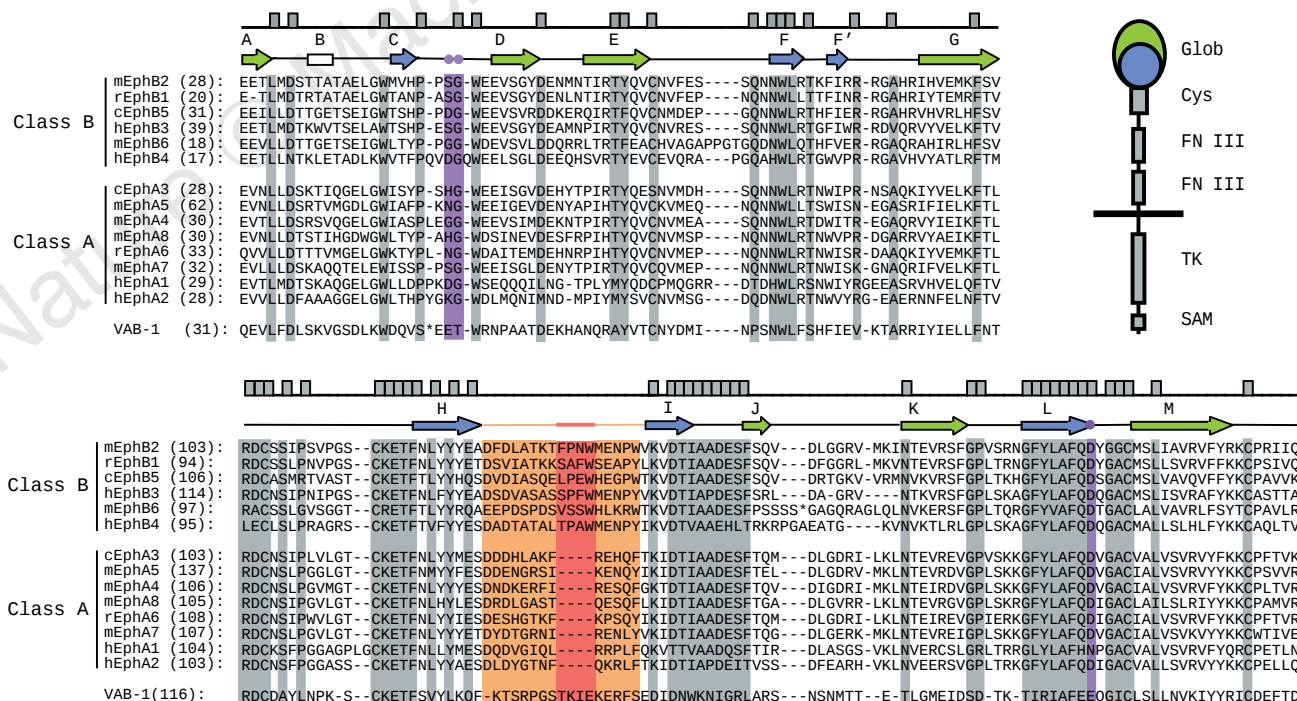


Figure 1 Alignment of amino-acid sequences of Eph globular domains. The sequences of the known Eph RTKs were aligned using the program DNASTar. The secondary-structure elements of the EphB2 crystal structure are shown using arrows for β -strands and a rectangle for the 3_{10} helix. The strands of the concave β -sheet are coloured blue and those in the convex β -sheet are coloured green. The locations of three different point mutations that affect the function of the *C. elegans* VAB-1 Eph receptor¹⁸ are indicated with purple dots and the Eph class-specificity loop is shown in orange, with the four residues by which the sub-

classes differ shown in red. The residues with the highest sequence conservation between RTKs (calculated with DNASTar) are indicated above the secondary structure and are also shaded in grey. Asterisks indicate the locations of a ten-amino-acid insertion in VAB-1 and a fifteen-amino-acid insertion in EphB6. The domain organization of the Eph receptors is shown at the right: Glob, globular domain; Cys, cysteine-rich linker; FN III, fibronectin type III motif; TK, tyrosine-kinase domain; SAM, sterile α -motif.

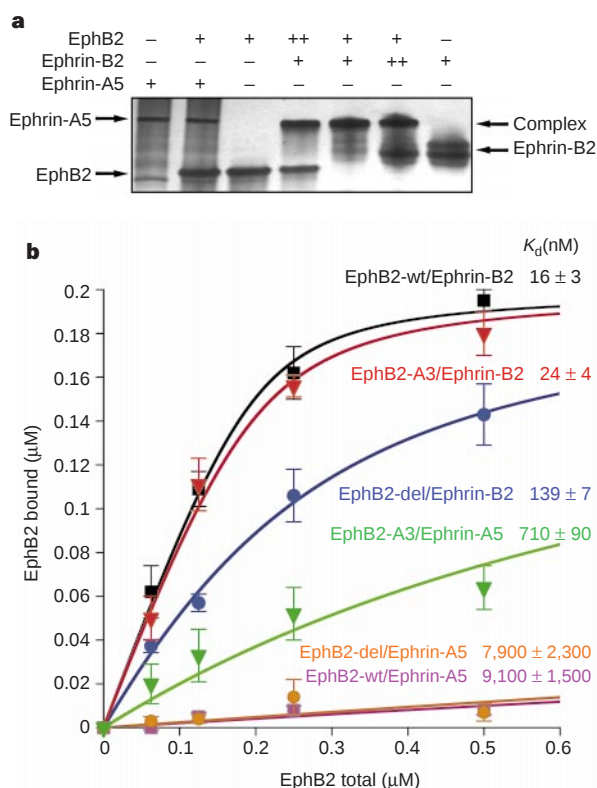


Figure 2 Eph-ephrin recognition. **a**, Gel mobility-shift experiment with the recombinant globular domain of EphB2 that was used in the structure determination and the recombinant extracellular domains of ephrin-B2 and ephrin-A5. The two bands in lanes with ephrin-B2 represent a monomer/dimer equilibrium that is also observed using size-exclusion chromatography. ++ indicates a twofold molar excess of the component. **b**, Binding of wild-type and mutant EphB2-receptor ligand-binding domains to class A and class B ephrins in Ni-affinity-bead pull-down experiments (see Methods). Data points represent the means of multiple measurements (typically $n = 3$) and the bars represent standard errors, $\sigma/(n - 1)$ (where σ is the standard deviation). The curves represent unweighted least-squares fits to a standard dissociation equilibrium equation. The calculated apparent dissociation constants are somewhat overestimated because the post-binding washing of the Ni-affinity-beads is not accounted for in the calculations.

Eph receptors can be divided into two subclasses on the basis of their preference to bind ligands that are tethered to the cell surface either by a glycosylphosphatidylinositol (GPI) anchor (A-ephrins) or by a single transmembrane segment (B-ephrins)¹². In general, EphA subclass receptors bind to the five known A-ephrins, whereas the EphB subclass receptors interact with the three known B-ephrins¹². Analysis of all currently available ligand-binding-domain sequences (Fig. 1) shows that the length of the loop connecting strands H and I may determine Eph-receptor subclasses. EphB members possess H-I loops that are 17 residues in length, whereas the H-I loops of EphA members are 4 residues shorter (orange/red loop in Fig. 4a). To understand better the role of this loop in ephrin recognition, we performed binding assays using two EphB2 mutants—one containing a deletion of the four extra class-B-specific amino acids (EphB2-del), and one in which the H-I loop of EphB2 was substituted with the corresponding loop of the A-class receptor EphA3 (EphB2-A3). As compared with the wild-type EphB2 protein, the EphB2-del mutant exhibited roughly ninefold reduced affinity for ephrin-B2, indicating that the H-I loop is directly involved in ligand binding. Furthermore, although both wild-type EphB2 and EphB2-del bind to ephrin-B2 only, the chimaeric EphB2-A3 receptor recognizes not only ephrin-B2 but also ephrin-A5, albeit with much lower affinity than a wild-type receptor for ephrin-A5, indicating that the loop confers some degree of class specificity. The ability of the EphB2-A3 mutant to retain full affinity for ephrin-B2 while also binding to ephrin-A5 indicates that the H-I loop may be particularly important for recognition of A-type ligands.

Other regions on the molecular surface of the globular domain of Eph-family receptors must also be involved in ligand recognition. In our crystal structure, the H-I loop packs against the middle of the concave β -sheet only 13 Å away from the locations of known signalling-defective mutations¹⁸ in the *Caenorhabditis elegans* Eph RTK VAB-1. The locations of these point mutations are shown in magenta in Fig. 4a. All characterized missense mutations in the extracellular region map to loops at the top of the β -sandwich. The *ju8* allele, which causes a strong loss-of-function phenotype¹⁸, contains a Glu → Lys substitution of a solvent-exposed residue (Ser 49 in EphB2) in the loop between strands C and D that will alter the local electrostatic potential. The intermediate alleles *e856* and *e699* contain a Thr → Ile substitution of a partially buried residue inside the C-D loop (Gly 50 in EphB2), and a Glu → Lys substitution of a residue buried inside the L-M loop (Asp 188 in EphB2), respectively. These mutations would probably alter the

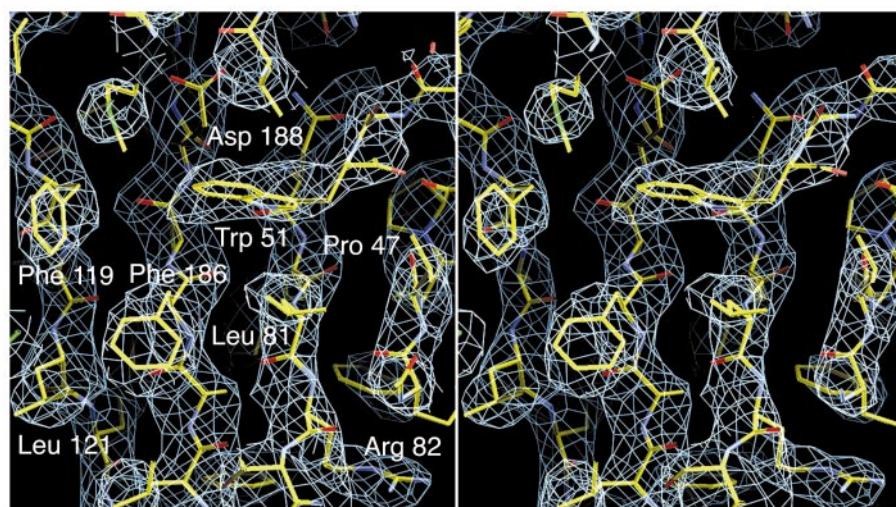


Figure 3 Stereo diagram of a representative region of the solvent-flattened experimental (MIR) electron-density map (contour level 1.2 σ) with the refined

EphB2 atomic model drawn as a stick figure.

conformation of the loops at the top of the β -sandwich (the exact nature of the perturbation is hard to predict, as EphB2 and VAB-1 have C-D loops of different lengths) but would probably not affect the overall folding of the domain. As Eph receptors also function as ligands in a bidirectional tyrosine-kinase signalling cascade^{4,13,14}, mutations in the ligand-binding region would affect signalling in

both directions, and indeed, the three described alleles produce stronger phenotypes than mutations that abolish the catalytic activity of VAB-1 (ref. 18).

In agreement with the direct evidence for the location of the ligand-binding region from the analyses of the VAB-1 and EphB2 H-I-loop mutants, the residues comprising the concave β -sheet

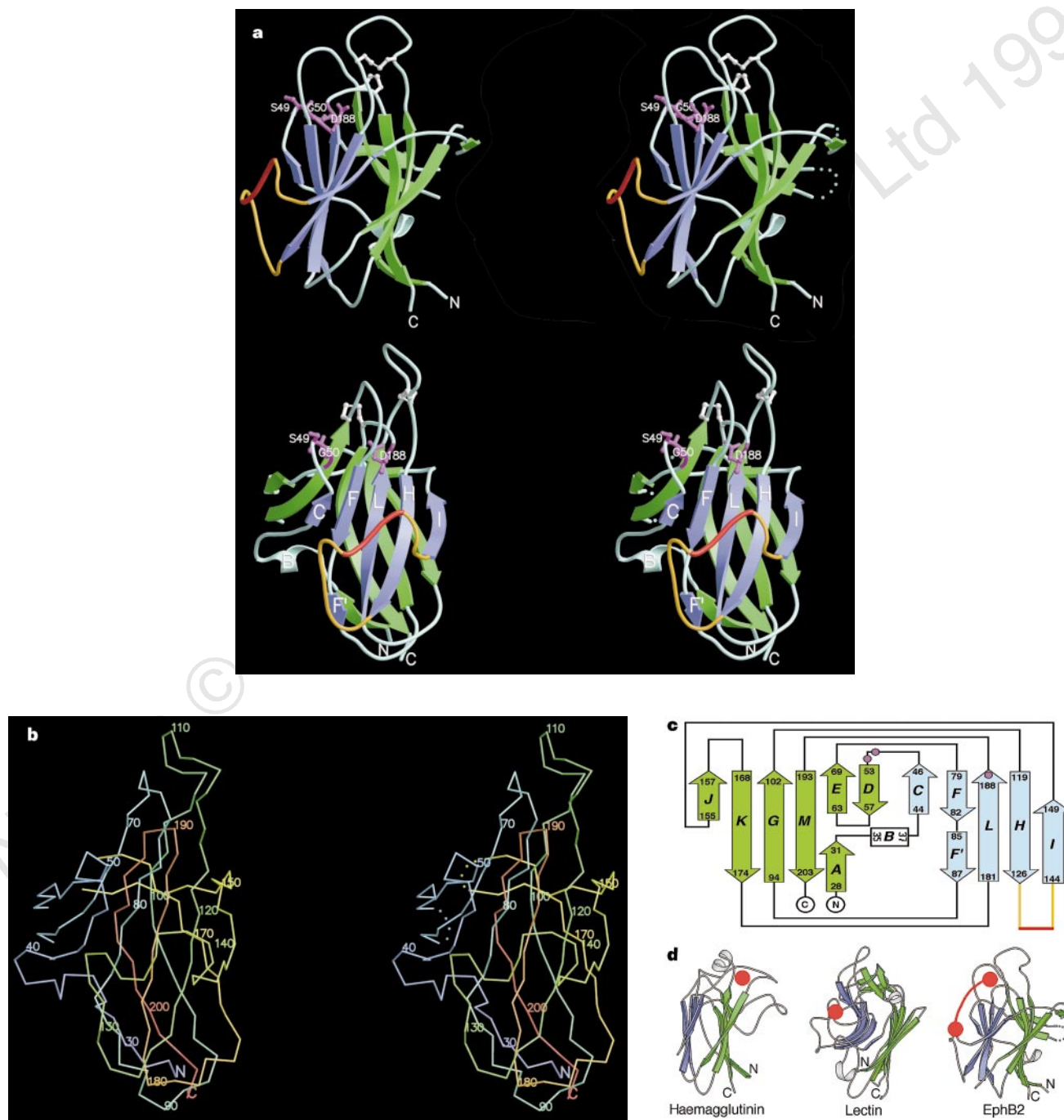


Figure 4 Structure of the N-terminal ligand-binding domain of EphB2. **a**, Two orthogonal stereoviews of the EphB2 ligand-binding domain. The N and C termini of the protein are labelled. The equivalent positions of VAB-1 point mutations are indicated as pink sidechain stick diagrams. The disulphide bridges are indicated as stick figures in white. The colour coding is as in Fig. 1. **b**, Stereoview of the C α trace of the EphB2 ligand-binding domain; the colour is graduated from the N terminus, blue, to the C terminus, red. Every tenth residue is numbered. **c**, Diagram of the secondary structure. Residues were assigned to secondary-structure elements according to criteria defined in ref. 30. The N and C termini of

the protein are labelled, as are the residues at the ends of each β -strand and of the short 3₁₀ helix. β -strands are shown as arrows and the helix as a rectangle. The colour coding is as in Fig. 1. **d**, Structures of two topological homologues of EphB2: left, the globular head of influenza virus haemagglutinin, with the location of the bound sialic acid shown in red²⁴; middle, *Griffonia simplicifolia* lectin, with the location of the bound human blood-group determinant shown in red²⁵; right, the globular domain of EphB2, with the proposed location for binding ephrin shown in red.

tend to be better conserved than those making up the convex β -sheet (Fig. 1). The secondary-structure elements that are most invariable are (Figs 1, 4) strand F, the G–H loop, strand I, the I–J loop, strand L and the L–M loop. Finally, the overall structural homology between the ligand-binding domain of EphB2 and the jellyroll carbohydrate-binding proteins indicates that they might use similar surface regions to bind their respective ligands. Several crystal and NMR structures reveal that the carbohydrates bind the receptors either at the top of the β -sandwich or at its concave face²⁰ and Fig. 4d shows two representative cases, haemagglutinin²⁴ and a legume-type lectin²⁵, for comparison with the ligand-binding domain of EphB2. Most structural studies of carbohydrate-binding proteins, however, involve low-affinity interactions ($K_d = 0.01$ – 0.1 mM) of mono-, di- or trisaccharides with a small region on the surface of the proteins, whereas a high-affinity interaction such as that between Eph receptors and their ligands would involve a larger protein–protein interface (Fig. 4d).

The structure of the globular domain of EphB2 is the first structure of a Eph RTK ligand-binding domain and reveals an unexpected molecular architecture. This is the first example, to our knowledge, of a *bona fide* protein-binding receptor with jellyroll folding topology. On the basis of the structure, published data and H–I-loop mutagenesis, we have located the surface region involved in ligand recognition, providing a starting point for further crystallographic, biochemical and genetic studies of Eph-receptor signalling. In particular, these data should aid directed, systematic analyses of ligand–receptor recognition to explain the high promiscuity within Eph subclasses but lack of interaction between subclasses. Furthermore, as we have already shown with the H–I loop, it should now be possible to design structure-based mutations that alter the affinity and specificity of receptor–ligand interactions to allow *in vivo* studies of how Eph receptor/ephrin expression boundaries and gradients direct the movement of cells and axons in the developing nervous system. □

Methods

Expression and crystallization. The sequence encoding the murine EphB2 globular domain (residues 28–210) was subcloned by the polymerase chain reaction (PCR) into a pET32 expression vector (Novagen), and overexpressed in *Escherichia coli* (strain AD494(DE3)) using the T7 RNA polymerase system. The protein was purified on a HiTrap Ni²⁺-chelating column (Pharmacia) and the histidine-containing N-terminal sequence was removed by thrombin proteolysis to yield EphB2(28–210) plus 36 N-terminal amino acids from the expression vector. A final anion-exchange chromatography step yielded material of homogeneity of 98% or greater. Mass spectrometry and DNA sequencing of the plasmid showed that the EphB2 ligand-binding domain used for crystallization was neither modified nor further proteolysed during expression and purification. Dynamic light scattering (Molecular Size Detector, Protein Solutions) showed that EphB2(28–210) was both monomeric and monodisperse at 1 mg ml^{−1}. The purified EphB2 protein was concentrated to 10 mg ml^{−1} in a buffer containing 10 mM KCl, 2 mM MgCl₂ and 10 mM HEPES, pH 7.2, and crystallized in a hanging drop by vapour diffusion against reservoir containing 200 mM ammonium sulphate, 15% PEG 4000 and 50 mM sodium acetate, pH 4.8. Crystals grew in the tetragonal space group $P4_212$ ($a = b = 55$ Å, $c = 159$ Å) with one molecule in the asymmetric unit. Heavy-atom derivatives were prepared by soaking the crystals in the heavy-metal reagents at concentrations of 1–5 mM for 2–3 days.

Mutagenesis and ligand-binding assays. Gel mobility-shift assays, used to confirm the biological activity of the recombinant proteins, were done by incubating the components at 0.2 mM concentration for 10 min at room temperature and separating them on 10–20% gradient non-denaturing polyacrylamide gels. EphB2 ligand-binding domain mutants (EphB2-del has a deletion of Phe 135 to Trp 138; EphB2-A3 includes a substitution of EphB2 residues Phe 128 to Trp 143 with EphA3 residues Asp 128 to Phe 139) were generated by PCR and confirmed by automated sequencing. The mutant proteins were expressed and purified as before. Both mutant proteins were properly folded, as judged by the fact that they were monomeric and

monodisperse at 0.01–0.2 mM concentrations and formed stable 1:1 complexes with ephrin-B2 in gel mobility-shift and size-exclusion chromatographic assays. Ephrin-B2 and ephrin-A5 were expressed and purified as before but with the histidine tag left intact.

The pull-down experiments were done as follows: the Eph proteins were incubated (at room temperature, for 30 min, in 1 ml total reaction volume, in a binding buffer containing 150 mM KCl, 2 mM MgCl₂, 20 mM imidazole, 20 mM HEPES, pH 8.2) at various concentrations with ephrins that were prebound to Ni-affinity beads (Pharmacia) (fixed ephrin concentration of 0.2 μ M). The beads were then isolated by centrifugation and washed with 1 ml binding buffer. The bound proteins were separated on 10–20% gradient polyacrylamide gels. The amount of bound EphB2 was estimated by the intensity of the Coomassie-blue staining (Fotodyne FotoAnalyst scanner and MacBas software), and corrected for background. Dissociation constants were determined from unweighted least-squares fits of the data to the standard equilibrium binding equation $K_d = [L][R]/[RL]$, where $[RL]$ is the concentration of the complex estimated from the amount of bound EphB2 $[R]$ is the concentration of the free EphB2 and $[L]$ is the concentration of free ephrin (in Fig. 2b, 'EphB2 bound' = $[RL]$, and 'EphB2 total' = $[RL] + [R]$; $[RL] + [L] = 0.2 \mu$ M). The estimated apparent dissociation constants are not corrected for complex dissociation during the washing step and, therefore, are used as relative and not absolute measures of the binding affinities of the EphB2–ephrin constructs.

Data collection and structure determination. Diffraction was measured at room temperature using a Rigaku RAXIS-IV imaging plate area detector. Oscillation photographs were integrated, scaled and merged using DENZO and SCALEPACK²⁶. Subsequent calculations were done with the CCP4 program suite²⁷. An initial MIR electron-density map was calculated with data between 15 Å and 3.2 Å (mean figure of merit 0.72) and improved by density modification (Fig. 3) which allowed unambiguous tracing and sequence assignment of the globular EphB2 domain using program O (ref. 28). Refinement of the model by conventional least-squares algorithm was done with XPLOR²⁹. The final refined model at 2.9 Å resolution (Table 1) has a free R value²⁹ of 31.4% and comprises 1,377 non-hydrogen atoms in 172 well-defined amino acids in the electron density map, and no water molecules. No electron density is observed for the six carboxy-terminal residues, for loop residues 59–61 and 160–163, and for the 34 N-terminal fusion residues introduced by the expression vector. Tightly restrained refinement of temperature factors was monitored throughout by the free- R -factor criterion²⁹. Stereochemical analysis of the refined model using PROCHECK (CCP4 suite) revealed main-chain and sidechain parameters better than, or within the typical range of, values for protein structures determined at 2.9 Å resolution (overall G -factor, -0.14). Only two EphB2 residues (Ala 39 and Ala 91) fell in the disallowed region of the Ramachandran plot.

Received 2 September; accepted 14 October 1998.

1. Eph Nomenclature Committee. Unified nomenclature for Eph family receptors and their ligands, the ephrins. *Cell* **90**, 403–440 (1997).
2. Flanagan, J. G. & Vanderhaeghe, P. The ephrins and Eph receptors in neural development. *Annu. Rev. Neurosci.* **21**, 309–345 (1998).
3. Drescher, U. *et al.* *In vitro* guidance of retinal ganglion cell axons by RAGS, a 25 kDa tectal protein related to ligands for Eph receptor tyrosine kinases. *Cell* **82**, 359–370 (1995).
4. Henkemeyer, M. *et al.* Nuk controls pathfinding of commissural axons in the mammalian central nervous system. *Cell* **86**, 35–46 (1996).
5. Orioli, D., Henkemeyer, M., Lemke, G., Klein, R. & Pawson, T. Sek4 and Nuk receptors cooperate in guidance of commissural axons and in palate formation. *EMBO J.* **15**, 6035–6049 (1996).
6. Park, S., Frisen, J. & Barbacid, M. Aberrant axonal projections in mice lacking EphA8 (Eek) tyrosine protein receptors. *EMBO J.* **16**, 3106–3114 (1997).
7. Winslow, J. W. *et al.* Cloning of AL-1, a ligand for an Eph-related tyrosine kinase receptor involved in axon bundle formation. *Neuron* **14**, 973–981 (1995).
8. Smith, A., Robinson, V., Patel, K. & Wilkinson, D. G. The EphA4 and EphB1 receptor tyrosine kinases and ephrin-B2 ligand regulate targeted migration of branchial neural crest cells. *Curr. Biol.* **7**, 561–570 (1997).
9. Xu, Q., Allard, G., Holder, N. & Wilkinson, D. G. Expression of truncated *Sek-1* receptor tyrosine kinase disrupts the segmental restriction of gene expression in the *Xenopus* and zebrafish hindbrain. *Development* **121**, 4005–4016 (1995).
10. Wang, H. U. & Anderson, D. J. Eph family transmembrane ligands can mediate repulsive guidance of trunk neural crest migration and motor axon outgrowth. *Neuron* **18**, 383–396 (1997).
11. Krull, C. E. *et al.* Interactions of Eph-related receptors and ligands confer rostrocaudal pattern to trunk neural crest migration. *Curr. Biol.* **7**, 571–580 (1997).
12. Gale, N. W. *et al.* Eph receptors and ligands comprise two major specificity subclasses and are reciprocally compartmentalized during embryogenesis. *Neuron* **17**, 9–19 (1996).
13. Holland, S. J. *et al.* Bi-directional signalling through the EPH-family receptor Nuk and its transmembrane ligands. *Nature* **383**, 722–725 (1996).
14. Bruckner, K., Pasquale, E. B. & Klein, R. Tyrosine phosphorylation of transmembrane ligands for Eph receptors. *Science* **275**, 1640–1643 (1997).

- Acknowledgements.** We thank I. Berry for technical support; P. D. Jeffrey for help with X-ray measurements; U. Drescher for ephrin-A5 DNA; and S. K. Burley, J. Goldberg, N. P. Pavletich and M. K. Rosen for useful suggestions. J.-P.H. is a Winston Foundation fellow. This work was supported by the DeWitt Wallace Fund and the V Foundation (D.B.N.), and by the Kent Waldrep National Paralysis Foundation for Basic Neuroscience Research (M.H.).

491

Geoscientists are not just rock stars

The geosciences are opening up to approaches that embrace multidisciplinary solutions to emerging and age-old problems. This should result in much interesting research, and challenges for scientists and funding agencies.

Brendan Horton

Billed as "The future of Earth sciences: the challenges and opportunities of multiple disciplines and diversity", the recent fiftieth anniversary of the American Geological Institute (AGI) in Washington, DC, was not only a birthday party, it was a chance to assess future needs and strategies. The need to make society aware of the importance of Earth science issues, and to pursue research that is relevant to society, were major themes of this one-day symposium. Other themes included multidisciplinary research, lateral thinking, and education and outreach.

Proponents of the multidisciplinary approach look set to get their opportunity to prove its value. Rita Colwell, director of the US National Science Foundation (NSF), recently called interdisciplinary research "nothing short of vital" and said she regarded this approach to science as "one of the major challenges for NSF in the coming years" (see *Nature* 396, 202; 1998).

For Mary Lou Zoback of the US Geological Survey, the grand challenge for the Earth sciences is to focus on themes rather than disciplines. She suggested several themes, including the need to recognize the signal of natural variability, to quantify the anthropogenic effect on climate and to look at the feedback loop between natural and perturbed systems. Zoback said a problem-based structure was needed at NSF rather than one based on traditional disciplines. Colwell seems to agree, having described "departmental structure as a block" to multidisciplinary research. Zoback hopes that a theme-orientated structure will focus more on the Earth's surface, integrating the study of the bio-, litho-, atmo- and hydrospheres



Shake, rattle and roll: IRIS's seismic monitor will be used in educational work (www.iris.edu).

to tackle today's most pressing problems.

Jonathan Price, state geologist for Nevada, says multidisciplinary work has been going on for some time in the applied geosciences, and this does not mean simply bringing together people who study reflection seismology, stratigraphy and plate tectonics. Several years ago, the Nevada Bureau of Mines and Geology put together an earthquake emergency scenario. It included a lot of geology and seismology, said Price, but also had input from people with expertise in engineering geology, civil and structural engineering, emergency management, medicine, building inspection, law enforcement, and disaster relief, as well as businesspeople and government officials. Price emphasized that the skills of traditional Earth scientists will continue to be important and stated that "realizing what one's limitations are will become increasingly important as the field becomes more multidisciplinary".

It is not always true that multidisciplinary study means a group activity, according to Stephen Stanley, a professor of Earth and

planetary sciences at Johns Hopkins University in Baltimore. His studies have taken him far afield into climatology, oceanography, physical anthropology, palaeontology, and other areas. He said: "I'm not an expert in all these fields, but I consult people who are — and I know how to read. Exciting cross-disciplinary ideas entail connections, and a group of individuals lacks the neuronal connections of a single brain. There is often much to be gained, not by collaborating with others, but by collaborating with yourself."

Lateral thinking may mean different things in different sectors. To Donald Paul, vice-president of technical and environmental affairs at Chevron, it might mean applying geoscience to being a consulting specialist, a product developer, a systems integrator or a marketing manager. According to Thomas Hamilton, chairman, president and chief executive of EEX Corporation, these concepts mean being comfortable with several disciplines and learning the language of integrated solutions. In today's global village, understanding political science, diplomacy and culture can be more important to a project than a specific technology. □

Problem solving for the whole Earth

Potter Wickware

New tools and specialisms in the Earth sciences have emerged to help manage humanity's predicament with the environment and its resources. Global positioning satellites and satellite altimetry open up new ranges of observation of the Earth in space and time.

Encouraging education and outreach initiatives

The AGI anniversary was an occasion to encourage its members to get involved in education and outreach (E&O), an area in need of attention in the geosciences. Wider exposure to Earth sciences should result from new US National Science Education Standards, which place Earth sciences on a par with biology, chemistry and physics in kindergarten through to twelfth-grade education.

The importance of E&O is not lost on Michelle Hall-Wallace of the

University of Arizona and Katherine Johnson, director of E&O at the Incorporated Research Institutions for Seismology (IRIS) in Washington, DC, both of whom are building programmes for educating non-specialists of all levels, with information and data available to the public (see Internet addresses).

Gregory van der Vink, director of planning at IRIS, has shown what can be done using public data in an educational setting. He and 12 students, from a class he teaches each fall at Princeton

University's Science and Technology Council and Department of Geoscience, published a paper on the increasing economic vulnerability to natural disasters in the United States (*Eos* 79, 534-537; 1998). The paper was presented in Washington, DC, after which a congressman immediately offered any of the 12 students a congressional staff position, according to van der Vink. The course was developed not for producers of scientific information

but for future consumers of that information in government or business.

AGI's Earth Week

<http://www.earthsciweek.org/>

Michelle Hall-Wallace page

<http://geo.arizona.edu/K-12/michellehall.html>

Incorporated Research Institutions for Seismology (IRIS)

<http://www.iris.washington.edu/>

Princeton Earth Physics Project

<http://lasker.princeton.edu/pepp.shtml>

Powerful modes of computer modelling allow experimental strategies involving large data sets to be devised and quickly refined. The Internet puts communities of scientists in instant communication with one another. Synchrotron radiation sources with powers many orders of magnitude greater than when X-rays were discovered allow views of unprecedented precision into materials.

Nevertheless, the problems of climate change, flooding, pollution and energy shortage are all too familiar in many parts of the world, while strategies for dealing with them often remain obscure. Research challenges at the most fundamental level await the efforts of geoscientists. For example, Miriam Kastner of the Scripps Institution of Oceanography in La Jolla, California, points out that we did not know of the existence of subsea vents until 1980. Since then we have discovered that the entire volume of the oceans cycles through them once every 5 million years, and through the subduction zones once every 200 million years.

"As we learn more about the interaction of the solid Earth and the hydrologic cycle we'll understand more about earthquakes, and possibly will be able to predict fine grain changes over time such as the Little Ice Age," she says. She adds that already we have found that some classic models of climate and geochemistry are wrong and must be revised. But this can only be done if scientists from disparate disciplines cooperate.

Geoscientists also have a vital contribution to make to human health in the area of environmental toxins and pollutants. There is an urgent need for more knowledge of dust particle size and chemistry, aerosol types, and the bioavailability of heavy metals, phosphate, arsenic and other substances. Here again, scientists must communicate not just across specialisms but across jurisdictions, because atmospheric pollutants are a global phenomenon. That the geoscience and medical communities have historically been disconnected from one another shows how "research is Balkanized," says Gordon Brown, of Stanford University's Synchrotron Radiation Laboratory. He suggests that perhaps the toxicologist is the logical interface through which this interaction will occur.

Charlie Alpers of the US Geological Survey in Sacramento, California, says that clear communication to outsiders about research problems and findings is also crucial. "We need to communicate clearly with the public, and persuade them and their representatives that what we are doing is worthwhile. Ultimately, public agencies and individual livelihoods depend on it." He says he will emphasize the human welfare aspect of what he does in a forthcoming short course in sulphate mineralogy, just as his agency accompanies many of the reports it publishes with clear fact sheets to make the work comprehensible to non-specialists, such as students.

Brown seconds this idea, saying that scientists ought to be able to make society care about their work. "Who cares about the distance between atoms?" he was recently asked by a political candidate in California. He gave an example of how the 0.4 Å difference between the radii of Cr^{6+} and Cr^{3+} quite possibly allows the latter to mimic phosphate in life systems, with potentially lethal results. He added that it is important from a public health perspective to accurately characterize the bioavailability of different metals.

Although the interdisciplinary approach is mandatory for effective science and policy, it can take getting used to, admits Joe Smith, of the Department of Geophysical Sciences at the University of Chicago. Fifteen years ago the American Geophysical Union's public affairs committee asked him to organize a scientific session on the 'nuclear winter' issue. Once involved, he was appalled to learn of the magnitude of the problem that science had helped to create, and it was at this time that he began to alter his view of science and the responsibilities of the scientist.

Especially now, when the Earth sciences are expanding so greatly in concepts and information density, he believes it is important for the Earth scientist to "think out the interconnections between what he or she does and the various pieces of the tapestry of human life". Even though it may be easier to work on idealized problems, he has come to realize that members of the US National Academy of Sciences like himself have a duty to use their expertise for human welfare.

The cost of future research may be considerable by the historical standards of geoscience. For example, the current third generation of synchrotron radiation sources costs about \$15 million per cluster of experimental workstations, each of which should be manned by two dozen scientists and engineers, says Smith. Available funds are too

scanty to serve the growing number of environmental geoscientists who require data from these instruments.

As relative latecomers to the science funding competition, Earth scientists face a struggle to win adequate support. Smith is preparing a presentation to the National Academy of Sciences to raise \$100 million over five years to support staff at synchrotron radiation centres, with matching amounts to support research, teaching and public outreach at universities and experimental stations. He says this sum is a tiny fraction of the money spent on nuclear weapons management, and is in the range of the cheapest rocket mission to an asteroid. □

Potter Wickware is a science writer in Oakland, California, USA.

e-mail: wickware@worldnet.att.net

A risk proposition

Brendan Horton

A geoscientist's career can veer off academic and government tracks into involvement in industry and even insurance. According to John Mutter, deputy director of Columbia University's Lamont-Doherty Earth Observatory, there are concerns to be aware of when too much of your funding comes from an industry-academic collaboration, especially for researchers intent on landing a rare academic post. But many scientists are diversifying their CVs to leave open employment options outside the academic world.

"More and more people are building in funding from the private sector," says Mutter, not only because it diversifies one's financial support, but also because people see the potential for a career in the private sector. While Lamont's partnerships have traditionally been dominated by the oil industry, "if we spoke two or three years from now, I'd [probably] be telling you about valuing climate information", says Mutter. With few exceptions, the industries most affected by climate have no research culture to enable them to understand it.

The winds of change that blew around Chris Barton, a research geologist with the US Geological Survey since 1985, may not have been hurricane force, but they were strong enough to make him a G. K. Gilbert Fellow at IBM with Benoit Mandelbrot, the 'father of fractals'. He followed this up with another fellowship at Lamont. These fellowships helped Barton's research to evolve from studying the physical phenomena of climate catastrophes to studying the magnitude of financial losses from these events, which in turn has affected the way reinsurance companies and the Federal Emergency Management Association evaluate risk. (Reinsurers insure major corporations and primary insurance providers.)

Further information on geoscience careers

GeoWeb Interactive: Jobs for GeoScience/Engineering

<http://www.ggrweb.com/job.html>

Earthworks

<http://ourworld.compuserve.com/homepages/eworks/>

USGS Main homepage

<http://www.usgs.gov/>

USGS Mineral Resources

<http://minerals.er.usgs.gov/>

NAS Geosciences

<http://www.nsf.gov/home/geo/start.htm>

AGI Careers in Geosciences

<http://www.agiweb.org/career/carehome.html>

Graduate School and the Job Market of the 1990s

http://www.agu.org/eos_elec/97117e.html

The Lamont-Doherty Earth Observatory

<http://www.ldeo.columbia.edu/>

The Columbia Earth Institute

<http://www.earthinstitute.columbia.edu/>

Alternative Careers in Science, edited by Cynthia Robbins-Roth

<http://www.amazon.com>

According to Barton, the statistics of the insurance industry have been well understood for almost 100 years. "It's the law of large numbers", which follows a gaussian curve in terms of losses. He explains that with these natural events the mathematics follow a power law or fractal. Until recently, he adds, the use of large-number statistics meant the insurance industry was tremendously undercapitalized for the actual risk.

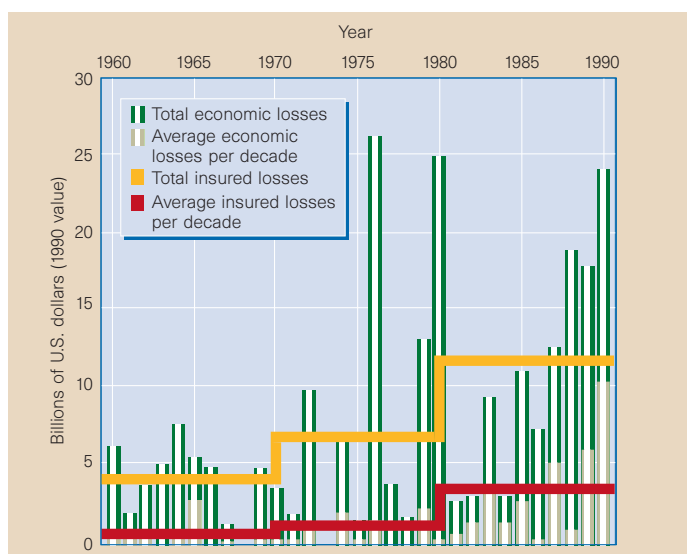
What has impressed Barton is that just as you can plot the strength of an event — be it the magnitude of an earthquake or the flood-discharge rate of a river — so you can plot dollar losses associated with each event (provided you have 50–100 years of data). "This behaviour was well understood for natural hazards, but no one had ever tried to do this for the property losses."

Hedge against catastrophe

Graciela Chichilnisky is a mathematician and an economist at Columbia University. She is also director of the Program on Information and Resources' Risk Center at the Columbia Earth Institute. Her interest in risk began in the early 1990s when she saw how catastrophic events such as hurricane Andrew had shaken up the insurance industry. Knowing that hurricane frequency is related to the El Niño Southern Oscillation (ENSO) and with her knowledge of chaotic, dynamical systems, Chichilnisky began to assess the risks from natural geophysical events, developing a theory of how these risks could be classified. She designed financial instruments that could effectively hedge against catastrophic events, which by their nature cannot be predicted. One result of her work is an instrument that she calls Catastrophe Bundles, a portion of which, the El Niño Index, is being commercialized by the Bermuda Stock Exchange in cooperation with Columbia University.

Chichilnisky's Risk Center operates in the grey area between publicly supported academic research and private interests, and allows the group's intellectual property to be commercialized. According to Chichilnisky, this set-up should provide better financial cover for the insurance industry, which may allow more companies to offer hurricane insurance at lower rates.

Some reinsurers, notably Swiss Re and



Since the 1960s, economic losses from natural disasters on a global scale have tripled, while insured losses have risen fivefold (after Berz, *Natural Hazards* 5, 95–102; 1992).

Munich Re, have built substantial in-house research groups that include scientists hired from academia. "I'm astonished to find myself working for a reinsurance company," says Jürg Trüb, head of the Atmospheric Perils Group at Swiss Re. "However, I quickly learned that the insurance industry and, in particular, reinsurers need highly specialized geoscientists to analyse the risks of natural catastrophes."

About one-third of his group's work is in research and development. "Our work is more 'purpose-orientated' and less in-depth research," he says. The group adapts and interprets the latest results in geoscience for the insurance industry. It must examine how weather forecasting and ENSO could affect underwriting, identify natural catastrophes that the company and its clients may be exposed to, and assess whether the group thoroughly understands and can manage these risks.

Beyond the academic sphere

Since being involved in writing "Climate science and insurance risk" (*Nature* 389, 225–227; 1997), Anthony Michaels has found that much of his work is in giving career advice. As director of the Wrigley Institute for Environmental Studies at the University of Southern California, Michaels heads a programme that will cross the university's departmental boundaries, linking science, engineering, medicine and business

to the technical assessment of risk, return and uncertainty in the insurance, banking and other risk-related industries.

Michaels says they aim to train students in the degree programme "so that they come out with something that has relevance beyond the academic sphere. This is a fertile area for climate and Earth scientists who have some modelling background, good maths skills and an interest in a business application of their background."

Wrigley offers study programmes not only in environment-related risks, such as hurricanes and earthquakes, but also in the pollution and health risks that affect health- and business-liability decisions on both regional and global scales.

The Sloan Foundation has provided funding to create two technical masters' degrees: one merges physics and business whereas the other is a masters' degree in environmental risk. The National Science Foundation is also funding PhD programmes in environmental science and engineering, one of which is in environmental risk assessment, says Michaels.

"The other part of the programme is the Risk Center that we're putting together." According to Michaels, this is still in the planning stage. The centre is structured to enhance the interdisciplinary connections within the university, so that faculty-level projects cross the spectrum of environmental risk, and involve the College of Letters, Arts and Sciences, the Marshall School of Business, the School of Engineering and the Medical School. These projects will look at natural systems, the risks to human activity and how they affect business decisions.

Turning back the flood

Cynthia Rosenzweig, from the US space agency NASA's Goddard Institute for Space Studies in New York, is involved in interdisciplinary programmes that have captured private and public interest, as well as their funding. "After years of people paying lip-

Further reading on climate risk, insurance and finance

A Calculus of Risk	http://www.scientificamerican.com/1998/0598issue/0598stix.html
Christopher Barton at USGS	http://coastal.er.usgs.gov/hurricane_forecast/barton2.html
The Risk Prediction Initiative, Bermuda Biological Station	http://www.bbsr.edu/agcihome/rpi/rpihome.html
Columbia's Program on Information Resources	http://www.earthinstitute.columbia.edu/units_partnerships/pir/index.html
The USC Wrigley Institute for Environmental Studies	http://wrigley.usc.edu/buss.shtml
US National Assessment	http://www.nacc.usgcrp.gov/
Munich Reinsurance	http://www.munichre.com/
Swiss Reinsurance Company	http://www.swissre.com/

service to interdisciplinary work, it is really happening now," says Rosenzweig. She leads the Metropolitan East Coast Region for the US National Assessment, which involves teams of researchers and stakeholders from such coastal zone sectors as infrastructure, water resources, human health and institutional decision-making. The National Assessment programme will study the region's vulnerability to climate change. The Metropolitan East Coast is crucial to the assessment, says Rosenzweig, because "we are focusing on the role and response of large cities such as New York".

Rosenzweig also heads an interdisciplinary science team involved in predicting the effect of El Niño on agricultural and marine systems as they relate to food supply. On the private side, she works with a consortium of four major agri-businesses, led by Pioneer Hy-Bred International, to evaluate how regions of interest around the world will be affected by climate change. □

Water and welfare: hydrology options

Potter Wickware

Water supply and quality issues worldwide give hydrologists a broad range of problem and opportunities to occupy them in the years ahead. Ismail Serageldin, vice-president for special programmes at the World Bank and chairman of the World Commission on Water for the 21st Century, says that existing technology can be applied for immediate benefit in developed and developing countries, for example in the reuse of minimally treated water, harvesting rainwater and preventing leaks. At the same time, researchers can work towards economic desalinization methods, discovery of deep aquifers by remote sensing, engineered crops for maximum water utilization in arid regions, and other advanced technologies.

Some areas of need are going completely unaddressed, he says. For example, more needs to be done to mitigate the effects of monsoon flooding and the associated problems of wastewater flushing and water quality. But little research is taking place anywhere in the world in this important area. Basic science is also needed to generate adequate baseline data to make accurate predictions and decisions. "Most environmental statistics at present are unreliable or incomplete, and projections are also bedevilled by the difficulty of modelling interactions among sectors," he says.

Indeed, "why must we rely on water as the means of sanitation, where water and human waste are dumped together, creating major problems and requiring treatments which most developing countries cannot afford?"

Serageldin asks. "Scientists should explore how to cut the link between water and sanitation, and thereby dramatically reduce water demand."

Hydrologists' knowledge also plays an increasingly important role in land use decisions, says William K. Michener, at the Joseph W. Jones Ecological Research Center in Newton, Georgia. In the United States, a trend in favour of managed flooding and restoration of wetlands can be seen in the Clinton administration's proposed Clean Water Initiative, which aims for a net increase of 100,000 wetland acres per year. This will be paid for with \$2.3 billion in new water-related funding over five years.

The great Midwestern floods of 1993 and 1995–96 taught the lesson that managed flooding is a more sensible strategy than the flood suppression attempts of the past, says Michener. He points out that the licences of many older flood control dams are coming up for renewal by the US Department of Energy. Each licence review provides an opportunity to establish minimum water flow levels and create flow regimes that more closely mimic natural ones. In some cases dams are dismantled altogether, leading to large-scale habitat restoration, as with the 161-year-old Edwards Dam on Maine's Kennebec River and similar proposals for stretches of Idaho's Snake River. In 1993, 160 licences affecting 262 dams on 105 rivers expired; half of these relicensing actions are still under way. Some 550 more dams are due for relicensing in the next 15 years.

Finally, hydrologists' expertise is called on for policy recommendations as efforts gain momentum to strengthen the Clean Water Act, reorganize the activities of the US Army Corps of Engineers, and reframe the federal flood insurance programme.

The after-effects of mining present a quite different sort of problem and opportunity for hydrologists. Kirk Nordstrom, a hydrologist with the US Geological Survey in Boulder, Colorado, describes the Iron Mountain Superfund site, where groundwater percolates through faults, saturating metal-sulphide formations and generating copious acid drainage. The lack of buffering potential in the surrounding matrix and the vigorous sulphide- and metal-oxidizing activity of acid-adapted soil bacteria produce a witches' brew of concentrated heavy metals in a negative pH solution upstream of fisheries, population centres and agricultural land in California's upper Sacramento Valley. Although the US Environmental Protection Agency has the situation in hand for the time being, a permanently engineered solution based on sound geological and hydrological knowledge has yet to be proposed.

Water supply and quality problems afflict many parts of the world, but the same sets of multidisciplinary technical, social and political skills are needed everywhere to effective-

ly manage them. D. M. Smith, an environmental plant physiologist with the Institute of Hydrology in Oxford, England, collaborates with scientists in Kenya to examine below groundwater interactions between trees and crops. Chiefly motivated by the problem of land degradation and desertification, Smith says he tries to develop systems of land use that balance exploitation of resources, including water, with conservation. This reduces to a problem of engineering combinations of trees and crops that fit a particular locality. But, after factoring in the social, cultural and economic issues, climate change and high rates of population growth, the problems become complex. "As a consequence," he says, "hydrologists working in this field will be most effective when they work together with, for example, agronomists, economists and anthropologists."

He also stresses that the work must be participatory, with farmers and scientists working together. When he worked in Niger, West Africa, Smith says he would sit under a shady tree with groups of farmers to decide on experiments that could be done in their fields. "The advantages of collaboration over prescription are plain," he says. "When something works, the farmers will have been instrumental in making it happen, will have assessed the risks inherent in changing their practices and will therefore be more likely to adopt the measure permanently."

Smith adds the qualification that, although the community-based approach is effective for adaptive research, it is less straightforward for process-oriented research. For example, the investigation of micrometeorological control of soil evaporation is difficult with participatory methods. This makes it harder to obtain funding for this type of research in developing countries.

Funding, of course, is critical to getting the job done at all. Just as water is maldistributed in the world, so is money for water-related research. The United Kingdom can support research by a couple of hundred scientists at its Institute of Hydrology. But a country such as Niger has scanty financial resources, although it has the scientists, trained locally and overseas. The solution, in Smith's view, is to "twin" research institutes in the north and south. Scientists in developing countries would be provided with funding, training and access to state-of-the-art techniques and instruments.

"I believe there is scope for twinning research institutes, rather as towns are twinned across Europe — between England and Romania, for example. Such a scheme would be particularly apt for environmental science, because scientists from the north and south face similar issues. Ideally, twinning would come with a long-term source of funding for training and reciprocal visits to promote collaboration, perhaps funded by the United Nations or World Bank," he says. □